



STIC Search Report

Biotech-Chem Library

STIC Database Tracking Number: 176372

TO: Konstantina Katcheves
Location: REM-2A60/2C70
Art Unit: 1636
Friday, January 20, 2006
Case Serial Number: 10/511327

From: Barb O'Bryen
Location: Biotech-Chem Library
Remsen 1a69
Phone: 571-272-2518

BOB
barbara.obryen@uspto.gov

Search Notes

RUSH

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2006 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 17, 2006, 18:06:53 ; Search time 278.233 Seconds
(without alignments)
11032.288 Million cell updates/sec

Title: US-10-511-327-1
Perfect score: 54
Sequence: 1 gatctgattgattgaacga.....ttcagcataaatgcgaaac 54

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0% -

Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl.*
1: gb_ba.*
2: gb_in.*
3: gb_env.*
4: gb_on.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pr.*
9: gb_ro.*
10: gb_sts.*
11: gb_sy.*
12: gb_un.*
13: gb_vi.*
14: gb_htg.*
15: gb_pl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------------|--------------------|
| 1 | 42.8 | 79.3 | 1319 | 11 AY781404 | AY781404 Synthetic |
| 2 | 42.8 | 79.3 | 1319 | 11 AY781404 | AY781404 Synthetic |
| 3 | 42.8 | 79.3 | 1351 | 11 AY781405 | AY781405 Synthetic |
| 4 | 42.8 | 79.3 | 1351 | 11 AY781405 | AY781405 Synthetic |
| 5 | 42.8 | 79.3 | 1524 | 11 AY781401 | AY781401 Synthetic |
| 6 | 42.8 | 79.3 | 1524 | 11 AY781401 | AY781401 Synthetic |
| 7 | 42.8 | 79.3 | 1808 | 11 AY781408 | AY781408 Synthetic |
| 8 | 42.8 | 79.3 | 1808 | 11 AY781408 | AY781408 Synthetic |
| 9 | 42.8 | 79.3 | 1937 | 11 AY781403 | AY781403 Synthetic |
| 10 | 42.8 | 79.3 | 1937 | 11 AY781403 | AY781403 Synthetic |
| 11 | 42.8 | 79.3 | 2037 | 11 AY781407 | AY781407 Synthetic |
| 12 | 42.8 | 79.3 | 2037 | 11 AY781407 | AY781407 Synthetic |
| 13 | 42.8 | 79.3 | 2062 | 11 AY781402 | AY781402 Synthetic |
| 14 | 42.8 | 79.3 | 2062 | 11 AY781402 | AY781402 Synthetic |
| 15 | 42.8 | 79.3 | 2162 | 11 AY781406 | AY781406 Synthetic |
| 16 | 42.8 | 79.3 | 2162 | 11 AY781406 | AY781406 Synthetic |
| 17 | 42.8 | 79.3 | 2413 | 11 AY781399 | AY781399 Synthetic |
| 18 | 42.8 | 79.3 | 2413 | 11 AY781399 | AY781399 Synthetic |

| | | | | | |
|----|------|------|-------|--------------|--------------------|
| 19 | 42.8 | 79.3 | 2933 | 11 AY781400 | AY781400 Synthetic |
| 20 | 42.8 | 79.3 | 2933 | 11 AY781400 | AY781400 Synthetic |
| 21 | 42.8 | 79.3 | 4791 | 11 AY738638 | AY738638 Cloning v |
| 22 | 42.8 | 79.3 | 4791 | 11 AY738638 | AY738638 Cloning v |
| 23 | 40.8 | 75.6 | 120 | 11 SYNGENE | M12560 Synthetic g |
| 24 | 39.8 | 73.7 | 324 | 11 SYNPMUEND | M15949 Synthetic B |
| 25 | 39.4 | 73.0 | 58 | 6 AR084420 | AR084420 Sequence |
| 26 | 39.4 | 73.0 | 71 | 6 AR084419 | AR084419 Sequence |
| 27 | 39.4 | 73.0 | 77 | 6 BD073246 | BD073246 In vitro |
| 28 | 39.4 | 73.0 | 77 | 6 AR353866 | AR353866 Sequence |
| 29 | 39.4 | 73.0 | 82 | 7 PMUBSR | M10177 Bacterioph |
| 30 | 39.4 | 73.0 | 83 | 6 AR084418 | AR084418 Sequence |
| 31 | 39.4 | 73.0 | 117 | 6 AR084417 | AR084417 Sequence |
| 32 | 39.4 | 73.0 | 150 | 7 PMUSEE1 | M10863 Bacterioph |
| 33 | 39.4 | 73.0 | 220 | 7 NCMU3R | X05582 Mu-derived |
| 34 | 39.4 | 73.0 | 220 | 7 PMUNE2 | M34920 Bacterioph |
| 35 | 39.4 | 73.0 | 240 | 1 ECOTRPA | M33723 Escherichia |
| 36 | 39.4 | 73.0 | 903 | 6 A02708 | A02708 pnm506 DNA |
| 37 | 39.4 | 73.0 | 4665 | 1 ECOPOLBDA | M35371 E.coli DNA |
| 38 | 39.4 | 73.0 | 7003 | 6 AR084426 | AR084426 Sequence |
| 39 | 39.4 | 73.0 | 15611 | 1 ECOPHNAQ | J05260 E.coli psid |
| 40 | 39.4 | 73.0 | 15611 | 6 AR229537 | AR229537 Sequence |
| 41 | 39.4 | 73.0 | 36717 | 7 AF083977 | AF083977 Bacteriop |
| 42 | 39.4 | 73.0 | 37199 | 11 AY860420 | AY860420 Cloning v |
| 43 | 37.4 | 69.3 | 240 | 1 ECOTRPA | M33723 Escherichia |
| 44 | 31.8 | 58.9 | 5755 | 1 ECOCYS | M32101 E.coli thio |
| 45 | 30.4 | 56.3 | 1925 | 1 ECOMALIXA | M28539 E.coli mali |

ALIGNMENTS

RESULT 1
AY781404
LOCUS AY781404 1319 bp DNA linear SYN 28-FEB-2005
DEFINITION Synthetic construct transposon mini-Mu transposon TnCR6, complete sequence.
ACCESSION AY781404
VERSION AY781404.1 GI:60171997
KEYWORDS synthetic construct
SOURCE synthetic construct
ORGANISM synthetic construct
REFERENCE 1 (bases 1 to 1319)
AUTHORS Zhang, C.
TITLE Direct Submission
JOURNAL Submitted (14-OCT-2004) Copyrat Pty Ltd, 27-31 Wright Street, Clayton, Victoria 3168, Australia
FEATURES
source
1..1319
/organism="synthetic construct"
/mol_type="other DNA"
/db_xref="taxon:32630"
repeat_region 1..1319
/transposon="mini-Mu transposon TnCR6"
300..959
/codon_start=1
/transl_table=1
/product="chloramphenicol acetyltransferase"
/protein_id="AAI14471.1"
/db_xref="GI:60171998"
translation="MEKKITGYTTVDISQWHRKEHFEAFQSAOQYNTQVTDITAF
LKTVKKNKHFYPAFIFILARLMAHPEFRMAKDGELVIDSVHPCTVPEHQTET
SSLSWEYDDDFROFLHLYSDVACVGENLAVFPKGFENMFVSANPMWVSTSDLVN
AMNDPAPVFTMCKYTTQDKVLMLPLAIQVHHAVCDGFHVRMLNELQQYCDWEQGG
A"

ORIGIN
Query Match 79.3%; Score 42.8; DB 11; Length 1319;
Best Local Similarity 87.0%; Pred. No. 0.00055;
Matches 47; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
QY 1 GATCTGATTGATGAACGAAACGCGAAGCGTTTCACGATNAATGCGAAAC 54


```

Oy 1 GATCTGATTGATGAACGAAACCGGAAACGGTTTCACGATAAATCGGAAAC 54
    |||||
Db 1350 GATCTGAAGCGCGCACGAAACCGGAAACGGTTTCACGATAAATCGGAAAC 1297

RESULT 5
AY781401
LOCUS Synthetic construct transposon mini-Mu transposon TnCR4, complete
DEFINITION
ACCESSION AY781401.1 GI:60171909
VERSION AY781401
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS Zhang,C., Kitsberg,D., Chy,H., Zhou,Q. and Morrison,J.R.
TITLE Transposon-mediated generation of targeting vectors for the
production of gene knockouts
JOURNAL (er) Nucleic Acids Res. 33 (3), E24 (2005) In press
PUBMED 15699181
AUTHORS
DIRECT SUBMISSION
TITLE Submitted (14-OCT-2004) CopyRat Pty Ltd, 27-31 Wright Street,
Clayton, Victoria 3168, Australia
LOCATION/Qualifiers
FEATURES
source
1..1524
/organism="synthetic construct"
/mol_type="other DNA"
/db_xref="taxon:32630"
repeat_region
1..1524
/transposon="mini-Mu transposon TnCR4"
CDS
168..1358
/codon_start=1
/transl_table=1
/product="tetracyclin resistance protein"
/protein_id="AAK14468.1"
/db_xref="GI:60171910"
/translating="MKSNNALIVLTGTVTLDAVGIGLVMPVLPGLLRDVIHSDSIASH
YGVLLALYALMOFLCAPVLGALSDRFRPVLASLIGATIDYAIMATTPVLTLYAG
RIVAGITGATGAVAGYADITDGEDRHRHFLMSACFGVMVAGPVAGGLGALSILH
APFLAAVLNGLNLLCGFLMQESHGERRPMPLEAFNPVSSFRWARGMTTVAALMTV
FFIMQLVQVPAALWIFGEDRFRWSATMIGLSLAVFGILHALAQAFVTGPTKRFG
KOATIGMAADALCYLLAFATRCWMAFPIMILLASGIGMPALQAMLSRQVDDHOG
QLOGSLAALTSITSITGLPVTIYAIASASTWGLAWIVGALYLVCLPALRRGAWSR
ATST"

ORIGIN
Query Match 79.3%; Score 42.8; DB 11; Length 1524;
Best Local Similarity 87.0%; Pred. No. 0.00054;
Matches 47; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Oy 1 GATCTGATTGATGAACGAAACCGGAAACGGTTTCACGATAAATCGGAAAC 54
    |||||
Db 1523 GATCTGAAGCGCGCACGAAACCGGAAACGGTTTCACGATAAATCGGAAAC 1470

RESULT 7
AY781408
LOCUS Synthetic construct transposon mini-Mu transposon TnCR1, complete
DEFINITION
ACCESSION AY781408
VERSION AY781408.1 GI:60172112
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS Zhang,C., Kitsberg,D., Chy,H., Zhou,Q. and Morrison,J.R.
TITLE Transposon-mediated generation of targeting vectors for the
production of gene knockouts
JOURNAL (er) Nucleic Acids Res. 33 (3), E24 (2005) In press
PUBMED 15699181
AUTHORS
DIRECT SUBMISSION
TITLE Submitted (14-OCT-2004) CopyRat Pty Ltd, 27-31 Wright Street,
Clayton, Victoria 3168, Australia
LOCATION/Qualifiers
FEATURES
source
1..1808
/organism="synthetic construct"
/mol_type="other DNA"
/db_xref="taxon:32630"
repeat_region
1..1808
/transposon="mini-Mu transposon TnCR1"
CDS
825..1484
/codon_start=1
/transl_table=1
/product="chloramphenicol acetyltransferase"

```

```

production of gene knockouts
(er) Nucleic Acids Res. 33 (3), E24 (2005) In press
PUBMED 15699181
AUTHORS
DIRECT SUBMISSION
TITLE Submitted (14-OCT-2004) CopyRat Pty Ltd, 27-31 Wright Street,
Clayton, Victoria 3168, Australia
LOCATION/Qualifiers
FEATURES
source
1..1524
/organism="synthetic construct"
/mol_type="other DNA"
/db_xref="taxon:32630"
repeat_region
1..1524
/transposon="mini-Mu transposon TnCR4"
CDS
168..1358
/codon_start=1
/transl_table=1
/product="tetracyclin resistance protein"
/protein_id="AAK14468.1"
/db_xref="GI:60171910"
/translating="MKSNNALIVLTGTVTLDAVGIGLVMPVLPGLLRDVIHSDSIASH
YGVLLALYALMOFLCAPVLGALSDRFRPVLASLIGATIDYAIMATTPVLTLYAG
RIVAGITGATGAVAGYADITDGEDRHRHFLMSACFGVMVAGPVAGGLGALSILH
APFLAAVLNGLNLLCGFLMQESHGERRPMPLEAFNPVSSFRWARGMTTVAALMTV
FFIMQLVQVPAALWIFGEDRFRWSATMIGLSLAVFGILHALAQAFVTGPTKRFG
KOATIGMAADALCYLLAFATRCWMAFPIMILLASGIGMPALQAMLSRQVDDHOG
QLOGSLAALTSITSITGLPVTIYAIASASTWGLAWIVGALYLVCLPALRRGAWSR
ATST"

ORIGIN
Query Match 79.3%; Score 42.8; DB 11; Length 1524;
Best Local Similarity 87.0%; Pred. No. 0.00054;
Matches 47; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Oy 1 GATCTGATTGATGAACGAAACCGGAAACGGTTTCACGATAAATCGGAAAC 54
    |||||
Db 1523 GATCTGAAGCGCGCACGAAACCGGAAACGGTTTCACGATAAATCGGAAAC 1470

RESULT 7
AY781408
LOCUS Synthetic construct transposon mini-Mu transposon TnCR1, complete
DEFINITION
ACCESSION AY781408
VERSION AY781408.1 GI:60172112
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS Zhang,C., Kitsberg,D., Chy,H., Zhou,Q. and Morrison,J.R.
TITLE Transposon-mediated generation of targeting vectors for the
production of gene knockouts
JOURNAL (er) Nucleic Acids Res. 33 (3), E24 (2005) In press
PUBMED 15699181
AUTHORS
DIRECT SUBMISSION
TITLE Submitted (14-OCT-2004) CopyRat Pty Ltd, 27-31 Wright Street,
Clayton, Victoria 3168, Australia
LOCATION/Qualifiers
FEATURES
source
1..1808
/organism="synthetic construct"
/mol_type="other DNA"
/db_xref="taxon:32630"
repeat_region
1..1808
/transposon="mini-Mu transposon TnCR1"
CDS
825..1484
/codon_start=1
/transl_table=1
/product="chloramphenicol acetyltransferase"

```


EHQGLAPAEPLPARKASMPDGEDLVVTHGDACLPNINWENGRPSGFDICGRLGVADRY
QDIALATRDIAEELGGWADRFVLVYGIAAPDSQRIAFYRLLEDFF"

ORIGIN

Query Match 79.3%; Score 42.8; DB 11; Length 1937;
Best Local Similarity 87.0%; Pred. No. 0.00053;
Matches 47; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Oy 1 GATCTGATTGATGAACGAAAAACGCGAAACGTTTCACGATAAATGCGAAAAAC 54
|||||
Db 1936 GATCTGAAGCGCGCACGAAAAACGCGAAACGTTTCACGATAAATGCGAAAAAC 1883

RESULT 11

AY781407
LOCUS AY781407
DEFINITION Synthetic construct transposon mini-Mu transposon TnCR8B, complete
sequence.
ACCESSION AY781407
VERSION AY781407.1 GI:60172085
KEYWORDS
SOURCE

ORGANISM

synthetic construct
other sequences; artificial sequences.
REFERENCE 1 (bases 1 to 2037)
AUTHORS Zhang, C., Kitsberg, D., Chy, H., Zhou, Q. and Morrison, J. R.
TITLE Transposon-mediated generation of targeting vectors for the
production of gene knockouts
JOURNAL (er) Nucleic Acids Res. 33 (3), E24 (2005) In press
PUBMED 15699181
REFERENCE 2 (bases 1 to 2037)
AUTHORS Zhang, C.
TITLE Direct Submission
JOURNAL Submitted (14-OCT-2004) CopyRat Pty Ltd, 27-31 Wright Street,
Clayton, Victoria 3168, Australia
FEATURES
source

repeat_region

1..2037
/organism="synthetic construct"
/mol_type="other DNA"
/db_xref="taxon:32630"
/transposon="mini-Mu transposon TnCR8B"
644..1438
/codon_start=1
/transl_table=1
/product="neomycin phosphotransferase"
/protein_id="AAI14474.1"
/db_xref="GI:60172086"
/translation="MIEQDGLHAGSPAAMVERLFGYDWAQOTICSDAAVRLSAQGR
PVLVFKTDLGALNELQDEARLSWLTATGVPCAAALVDVVTAGRDWLLGLGEVPGQDL
LSSHAPAEKVSIMADAMRLHTLDPATCFDHOAKHRIERARTRMEAGLVDDDDLE
EHQGLAPAEPLPARKASMPDGEDLVVTHGDACLPNINWENGRPSGFDICGRLGVADRY
QDIALATRDIAEELGGWADRFVLVYGIAAPDSQRIAFYRLLEDFF"

CDS

Query Match 79.3%; Score 42.8; DB 11; Length 2037;
Best Local Similarity 87.0%; Pred. No. 0.00052;
Matches 47; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Oy 1 GATCTGATTGATGAACGAAAAACGCGAAACGTTTCACGATAAATGCGAAAAAC 54
|||||
Db 2 GATCTGAAGCGCGCACGAAAAACGCGAAACGTTTCACGATAAATGCGAAAAAC 55

ORIGIN

Query Match 79.3%; Score 42.8; DB 11; Length 2037;
Best Local Similarity 87.0%; Pred. No. 0.00052;
Matches 47; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Oy 1 GATCTGATTGATGAACGAAAAACGCGAAACGTTTCACGATAAATGCGAAAAAC 54
|||||
Db 2 GATCTGAAGCGCGCACGAAAAACGCGAAACGTTTCACGATAAATGCGAAAAAC 55

RESULT 12

AY781407/c
LOCUS AY781407
DEFINITION Synthetic construct transposon mini-Mu transposon TnCR8B, complete
sequence.
ACCESSION AY781407
VERSION AY781407.1 GI:60172085
KEYWORDS
SOURCE

repeat_region

1..2062
/organism="synthetic construct"
/mol_type="other DNA"
/db_xref="taxon:32630"
/transposon="mini-Mu transposon TnCR5A"
732..1526
/codon_start=1
/transl_table=1
/product="neomycin phosphotransferase"
/protein_id="AAI14469.1"
/db_xref="GI:60171935"
/translation="MIEQDGLHAGSPAAMVERLFGYDWAQOTICSDAAVRLSAQGR
PVLVFKTDLGALNELQDEARLSWLTATGVPCAAALVDVVTAGRDWLLGLGEVPGQDL

ORGANISM

synthetic construct
other sequences; artificial sequences.
REFERENCE 1 (bases 1 to 2037)
AUTHORS Zhang, C., Kitsberg, D., Chy, H., Zhou, Q. and Morrison, J. R.
TITLE Transposon-mediated generation of targeting vectors for the
production of gene knockouts
JOURNAL (er) Nucleic Acids Res. 33 (3), E24 (2005) In press
PUBMED 15699181
REFERENCE 2 (bases 1 to 2037)
AUTHORS Zhang, C.
TITLE Direct Submission
JOURNAL Submitted (14-OCT-2004) CopyRat Pty Ltd, 27-31 Wright Street,
Clayton, Victoria 3168, Australia
FEATURES
source

repeat_region

1..2037
/organism="synthetic construct"
/mol_type="other DNA"
/db_xref="taxon:32630"
/transposon="mini-Mu transposon TnCR8B"
644..1438
/codon_start=1
/transl_table=1
/product="neomycin phosphotransferase"
/protein_id="AAI14474.1"
/db_xref="GI:60172086"
/translation="MIEQDGLHAGSPAAMVERLFGYDWAQOTICSDAAVRLSAQGR
PVLVFKTDLGALNELQDEARLSWLTATGVPCAAALVDVVTAGRDWLLGLGEVPGQDL
LSSHAPAEKVSIMADAMRLHTLDPATCFDHOAKHRIERARTRMEAGLVDDDDLE
EHQGLAPAEPLPARKASMPDGEDLVVTHGDACLPNINWENGRPSGFDICGRLGVADRY
QDIALATRDIAEELGGWADRFVLVYGIAAPDSQRIAFYRLLEDFF"

CDS

Query Match 79.3%; Score 42.8; DB 11; Length 2037;
Best Local Similarity 87.0%; Pred. No. 0.00052;
Matches 47; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Oy 1 GATCTGATTGATGAACGAAAAACGCGAAACGTTTCACGATAAATGCGAAAAAC 54
|||||
Db 2036 GATCTGAAGCGCGCACGAAAAACGCGAAACGTTTCACGATAAATGCGAAAAAC 1983

ORIGIN

Query Match 79.3%; Score 42.8; DB 11; Length 2037;
Best Local Similarity 87.0%; Pred. No. 0.00052;
Matches 47; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

RESULT 13
AY781402
LOCUS AY781402
DEFINITION Synthetic construct transposon mini-Mu transposon TnCR5A, complete
sequence.
ACCESSION AY781402
VERSION AY781402.1 GI:60171934
KEYWORDS
SOURCE
ORGANISM
synthetic construct
other sequences; artificial sequences.
REFERENCE 1 (bases 1 to 2062)
AUTHORS Zhang, C.
TITLE Direct Submission
JOURNAL Submitted (14-OCT-2004) CopyRat Pty Ltd, 27-31 Wright Street,
Clayton, Victoria 3168, Australia
FEATURES
source

repeat_region

1..2062
/organism="synthetic construct"
/mol_type="other DNA"
/db_xref="taxon:32630"
/transposon="mini-Mu transposon TnCR5A"
732..1526
/codon_start=1
/transl_table=1
/product="neomycin phosphotransferase"
/protein_id="AAI14469.1"
/db_xref="GI:60171935"
/translation="MIEQDGLHAGSPAAMVERLFGYDWAQOTICSDAAVRLSAQGR
PVLVFKTDLGALNELQDEARLSWLTATGVPCAAALVDVVTAGRDWLLGLGEVPGQDL

CDS

Query Match 79.3%; Score 42.8; DB 11; Length 2037;
Best Local Similarity 87.0%; Pred. No. 0.00052;
Matches 47; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

RESULT 13
AY781402
LOCUS AY781402
DEFINITION Synthetic construct transposon mini-Mu transposon TnCR5A, complete
sequence.
ACCESSION AY781402
VERSION AY781402.1 GI:60171934
KEYWORDS
SOURCE
ORGANISM
synthetic construct
other sequences; artificial sequences.
REFERENCE 1 (bases 1 to 2062)
AUTHORS Zhang, C.
TITLE Direct Submission
JOURNAL Submitted (14-OCT-2004) CopyRat Pty Ltd, 27-31 Wright Street,
Clayton, Victoria 3168, Australia
FEATURES
source

repeat_region

1..2062
/organism="synthetic construct"
/mol_type="other DNA"
/db_xref="taxon:32630"
/transposon="mini-Mu transposon TnCR5A"
732..1526
/codon_start=1
/transl_table=1
/product="neomycin phosphotransferase"
/protein_id="AAI14469.1"
/db_xref="GI:60171935"
/translation="MIEQDGLHAGSPAAMVERLFGYDWAQOTICSDAAVRLSAQGR
PVLVFKTDLGALNELQDEARLSWLTATGVPCAAALVDVVTAGRDWLLGLGEVPGQDL

CDS

Query Match 79.3%; Score 42.8; DB 11; Length 2037;
Best Local Similarity 87.0%; Pred. No. 0.00052;
Matches 47; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

RESULT 13
AY781402
LOCUS AY781402
DEFINITION Synthetic construct transposon mini-Mu transposon TnCR5A, complete
sequence.
ACCESSION AY781402
VERSION AY781402.1 GI:60171934
KEYWORDS
SOURCE
ORGANISM
synthetic construct
other sequences; artificial sequences.
REFERENCE 1 (bases 1 to 2062)
AUTHORS Zhang, C.
TITLE Direct Submission
JOURNAL Submitted (14-OCT-2004) CopyRat Pty Ltd, 27-31 Wright Street,
Clayton, Victoria 3168, Australia
FEATURES
source

repeat_region

1..2062
/organism="synthetic construct"
/mol_type="other DNA"
/db_xref="taxon:32630"
/transposon="mini-Mu transposon TnCR5A"
732..1526
/codon_start=1
/transl_table=1
/product="neomycin phosphotransferase"
/protein_id="AAI14469.1"
/db_xref="GI:60171935"
/translation="MIEQDGLHAGSPAAMVERLFGYDWAQOTICSDAAVRLSAQGR
PVLVFKTDLGALNELQDEARLSWLTATGVPCAAALVDVVTAGRDWLLGLGEVPGQDL

LSSLAPAEKVSIMADMRRLHTLDPATCPDHOAKHRIERARTMEAGLVDDQLDE
EHQGLAPAEFLPAELKARMPDGEDLVVTHGDACLPIINVENGPSGFTDCGRLGVADRY
QDIALATRDIAEELGGEWADRFVLVYGIAAPDSQRIAFYRLDDEFF"

ORIGIN

Query Match 79.3%; Score 42.8; DB 11; Length 2062;
Best Local Similarity 87.0%; Pred. No. 0.00052;
Matches 47; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 1 GATCTGATTGATTGAACGAAACCGGAAACGGTTTCACGATAAATCGGAAAC 54
|||||
Db 2 GATCTGAAGCGCGCACGAAACCGGAAACGGTTTCACGATAAATCGGAAAC 55
|||||

RESULT 14
AY781402/c
LOCUS AY781402 2062 bp DNA linear SYN 28-FEB-2005
DEFINITION Synthetic construct transposon mini-Mu transposon TnCR5A, complete
sequence.
ACCESSION AY781402
VERSION AY781402.1 GI:60171934
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
LOCATION/Qualifiers
FEATURES
source
1..2062
/organism="synthetic construct"
/mol_type="other DNA"
/db_xref="taxon:32630"
repeat_region 1..2062
/transposon="mini-Mu transposon TnCR5A"
CDS
732..1526
/codon_start=1
/transl_table=1
/product="neomycin phosphotransferase"
/protein_id="AAI14469.1"
/db_xref="GI:60171935"
/translation="MIEQDGLHAGSPAANVERLFGYDWAQQTIGCSDAAVFRLSAQGR
PVLVKTDLGALNELQDEARLSWLATTPCAAVLDVVTEAGRDWLLIGEVPGQDL
LSSLAPAEKVSIMADMRRLHTLDPATCPDHOAKHRIERARTMEAGLVDDQLDE
EHQGLAPAEFLPAELKARMPDGEDLVVTHGDACLPIINVENGPSGFTDCGRLGVADRY
QDIALATRDIAEELGGEWADRFVLVYGIAAPDSQRIAFYRLDDEFF"

Direct Submission
Submitted (14-OCT-2004) CopyRat Pty Ltd, 27-31 Wright Street,
Clayton, Victoria 3168, Australia
Location/Qualifiers
1..2062
/organism="synthetic construct"
/mol_type="other DNA"
/db_xref="taxon:32630"
repeat_region 1..2062
/transposon="mini-Mu transposon TnCR5A"
CDS
732..1526
/codon_start=1
/transl_table=1
/product="neomycin phosphotransferase"
/protein_id="AAI14469.1"
/db_xref="GI:60171935"
/translation="MIEQDGLHAGSPAANVERLFGYDWAQQTIGCSDAAVFRLSAQGR
PVLVKTDLGALNELQDEARLSWLATTPCAAVLDVVTEAGRDWLLIGEVPGQDL
LSSLAPAEKVSIMADMRRLHTLDPATCPDHOAKHRIERARTMEAGLVDDQLDE
EHQGLAPAEFLPAELKARMPDGEDLVVTHGDACLPIINVENGPSGFTDCGRLGVADRY
QDIALATRDIAEELGGEWADRFVLVYGIAAPDSQRIAFYRLDDEFF"

Query Match 79.3%; Score 42.8; DB 11; Length 2062;
Best Local Similarity 87.0%; Pred. No. 0.00052;
Matches 47; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 1 GATCTGATTGATTGAACGAAACCGGAAACGGTTTCACGATAAATCGGAAAC 54
|||||
Db 2061 GATCTGAAGCGCGCACGAAACCGGAAACGGTTTCACGATAAATCGGAAAC 2008
|||||

ORIGIN

Query Match 79.3%; Score 42.8; DB 11; Length 2062;
Best Local Similarity 87.0%; Pred. No. 0.00052;
Matches 47; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 1 GATCTGATTGATTGAACGAAACCGGAAACGGTTTCACGATAAATCGGAAAC 54
|||||
Db 2061 GATCTGAAGCGCGCACGAAACCGGAAACGGTTTCACGATAAATCGGAAAC 2008
|||||

RESULT 15
AY781406
LOCUS AY781406 2162 bp DNA linear SYN 28-FEB-2005
DEFINITION Synthetic construct transposon mini-Mu transposon TnCR8A, complete
sequence.
ACCESSION AY781406
VERSION AY781406.1 GI:60172057
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
Transposon-mediated generation of targeting vectors for the

production of gene knockouts
(er) Nucleic Acids Res. 33 (3), E24 (2005) In press
15699181
REFERENCE 2 (bases 1 to 2162)
AUTHORS
TITLE
JOURNAL
Clayton, Victoria 3168, Australia
Location/Qualifiers
1..2162
/organism="synthetic construct"
/mol_type="other DNA"
/db_xref="taxon:32630"
repeat_region 1..2162
/transposon="mini-Mu transposon TnCR8A"
CDS
769..1563
/codon_start=1
/transl_table=1
/product="neomycin phosphotransferase"
/protein_id="AAI14473.1"
/db_xref="GI:60172058"
/translation="MIEQDGLHAGSPAANVERLFGYDWAQQTIGCSDAAVFRLSAQGR
PVLVKTDLGALNELQDEARLSWLATTPCAAVLDVVTEAGRDWLLIGEVPGQDL
LSSLAPAEKVSIMADMRRLHTLDPATCPDHOAKHRIERARTMEAGLVDDQLDE
EHQGLAPAEFLPAELKARMPDGEDLVVTHGDACLPIINVENGPSGFTDCGRLGVADRY
QDIALATRDIAEELGGEWADRFVLVYGIAAPDSQRIAFYRLDDEFF"

Query Match 79.3%; Score 42.8; DB 11; Length 2162;
Best Local Similarity 87.0%; Pred. No. 0.00052;
Matches 47; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 1 GATCTGATTGATTGAACGAAACCGGAAACGGTTTCACGATAAATCGGAAAC 54
|||||
Db 2 GATCTGAAGCGCGCACGAAACCGGAAACGGTTTCACGATAAATCGGAAAC 55
|||||

Search completed: January 18, 2006, 04:25:12
Job time : 280.233 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2006 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 17, 2006, 17:02:23 ; Search time 33.5794 Seconds
(without alignments)
10717.687 Million cell updates/sec

Title: US-10-511-327-1
Perfect score: 54
Sequence: 1 gatctgattgattgaacgaa.....ttcacgataaatgcgaaac 54

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_21.*
1: Geneseq1980s.*
2: Geneseq1990s.*
3: Geneseq2000s.*
4: Geneseq2001as.*
5: Geneseq2001bs.*
6: Geneseq2002as.*
7: Geneseq2002bs.*
8: Geneseq2003as.*
9: Geneseq2003bs.*
10: Geneseq2003cs.*
11: Geneseq2003ds.*
12: Geneseq2004as.*
13: Geneseq2004bs.*
14: Geneseq2005s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------------|---------------------|
| 1 | 54 | 100.0 | 54 | 10 ACF58168 | Acf58168 Cat-Mu tr |
| 2 | 54 | 100.0 | 1254 | 10 ACF58169 | Acf58169 Modified |
| 3 | 54 | 100.0 | 1254 | 10 ACF58169 | Acf58169 Modified |
| 4 | 50 | 92.6 | 50 | 10 ACF58172 | Acf58172 Cat-Mu tr |
| 5 | 42.8 | 79.3 | 54 | 4 AAD21279 | Aad21279 Precut tr |
| 6 | 42.8 | 79.3 | 54 | 10 ACF58170 | Acf58170 Cat-Mu tr |
| 7 | 42.8 | 79.3 | 82 | 10 ACC80745 | Acc80745 Transposo |
| 8 | 42.8 | 79.3 | 83 | 10 ACC80741 | Acc80741 Transposo |
| 9 | 42.8 | 79.3 | 84 | 10 ACC80738 | Acc80738 Transposo |
| 10 | 42.8 | 79.3 | 86 | 10 ACC80742 | Acc80742 Transposo |
| 11 | 42.8 | 79.3 | 89 | 10 ACC80753 | Acc80753 Transposo |
| 12 | 41.2 | 76.3 | 54 | 10 ACF58171 | Acf58171 Cat-Mu (No |
| 13 | 39.4 | 73.0 | 51 | 13 ADS13821 | Ads13821 Mu-end DN |
| 14 | 39.4 | 73.0 | 56 | 13 ADS13820 | Ads13820 Mu-end DN |
| 15 | 39.4 | 73.0 | 58 | 2 AAZ28883 | Aaz28883 Mini-Mu t |
| 16 | 39.4 | 73.0 | 71 | 2 AAZ28882 | Aaz28882 Mini-Mu t |
| 17 | 39.4 | 73.0 | 83 | 2 AAZ28881 | Aaz28881 Mini-Mu t |
| 18 | 39.4 | 73.0 | 90 | 13 ADRI4827 | Adri4827 PCR prime |
| 19 | 39.4 | 73.0 | 117 | 2 AAZ28880 | Aaz28880 Mini-Mu t |

| | | | | | |
|----|------|------|--------|-------------|--------------------|
| 20 | 39.4 | 73.0 | 903 | 1 AAN80102 | Aan80102 EcoRI/Bam |
| 21 | 39.4 | 73.0 | 7003 | 2 AAZ28877 | Aaz28877 Transpos |
| 22 | 39.4 | 73.0 | 15611 | 3 RAD01008 | Rad01008 Escherich |
| 23 | 28 | 51.9 | 30 | 4 AAD21278 | Aad21278 Bacteriop |
| 24 | 25.8 | 47.8 | 1874 | 6 AAS98204 | Aas98204 DNA encod |
| 25 | 25.8 | 47.8 | 2229 | 7 ADS73196 | Ads73196 Human kid |
| 26 | 25.8 | 47.8 | 2229 | 7 ADW42050 | Adw42050 CDNA elev |
| 27 | 25.8 | 47.8 | 6136 | 6 ABL70342 | AbL70342 Chemical |
| 28 | 25.8 | 47.8 | 6136 | 6 RAS61297 | Ras61297 Human gen |
| 29 | 25.8 | 47.8 | 6730 | 11 ADL22564 | Adl22564 Human dis |
| 30 | 25.8 | 47.8 | 17687 | 4 AAK71665 | Aak71665 Human imm |
| 31 | 25.8 | 47.8 | 17687 | 4 AAK64966 | Aak64966 Human imm |
| 32 | 25.8 | 47.8 | 17979 | 4 AAK71664 | Aak71664 Human imm |
| 33 | 25.8 | 47.8 | 17979 | 4 AAK64964 | Aak64964 Human imm |
| 34 | 25 | 46.3 | 34 | 14 AEB45577 | Aeb45577 SigA2 wit |
| 35 | 25 | 46.3 | 34 | 14 AEB48763 | Aeb48763 Beta-lact |
| 36 | 24.4 | 45.2 | 2000 | 11 ACL38625 | Acl38625 Rice stre |
| 37 | 24.2 | 44.8 | 1995 | 10 ADE63262 | Ade63262 Rat gene |
| 38 | 24.2 | 44.8 | 1995 | 10 ADD46332 | Add46332 Rat gene |
| 39 | 24 | 44.4 | 2099 | 6 ABQ69021 | Abq69021 Listeria |
| 40 | 24 | 44.4 | 7868 | 6 ABQ71053 | Abq71053 Listeria |
| 41 | 24 | 44.4 | 110000 | 6 ABQ69245 | AbQ69245 Listeria |
| 42 | 24 | 44.4 | 110000 | 6 ABQ67197 | AbQ67197 Listeria |
| 43 | 23.8 | 44.1 | 2022 | 13 ADR85671 | Adr85671 Aspergill |
| 44 | 23.8 | 44.1 | 2022 | 13 ADR85084 | Adr85084 Aspergill |
| 45 | 23.8 | 44.1 | 4861 | 4 ABL29670 | AbL29670 Drosophi |

ALIGNMENTS

RESULT 1
ACF58168
ID ACF58168 standard; DNA; 54 BP.
XX
AC ACF58168;
XX
DT 15-JAN-2004 (first entry)
XX
DE Cat-Mu transposon modified end fragment.
XX
KW Transposon; genetic engineering; transposase; Cat-Mu; ds.
XX
OS Synthetic.
OS Bacteriophage mu.
XX
FH Key Location/Qualifiers
FT misc_feature 1..4
FT /note= a
FT /note= "the 5' end of this strand overhangs the 3' end of
FT the complementary strand"
XX
XX WO2003087370-A1.

XX
PD 23-OCT-2003.
XX
PF 14-APR-2003; 2003WO-FI000285.
XX
PR 18-APR-2002; 2002FI-00000746.
XX
XX (FINN-) FINNZYMES OY.
XX
XX Savilahti H, Tieaho V;
XX
XX WPI; 2003-845329/78.

XX
XX New transposon nucleic acid comprising a genetically engineered
XX translation stop signal within a transposon end sequence recognized by a
XX transposase useful for producing deletion derivatives of polypeptide.
XX
XX Claim 6; Fig 2; Opp; English.
XX
XX The invention relates to a transposon nucleic acid comprising a
XX


```

AC ACF58172;
XX
DT 15-JAN-2004 (first entry)
XX
DE Cat-Mu transposon modified end fragment.
XX
KW Transposon; genetic engineering; transposase; Cat-Mu; ds.
XX
OS Synthetic.
OS Bacteriophage mu.
XX
PN WO2003087370-A1.
XX
PD 23-OCT-2003.
XX
PF 14-APR-2003; 2003WO-FI000285.
XX
PR 18-APR-2002; 2002FI-00000746.
XX
PA (FINN-) FINNZYMES OY.
XX
PI Savilahti H, Tieaho V;
XX
DR WPI; 2003-845329/78.
XX
PT New transposon nucleic acid comprising a genetically engineered
PT translation stop signal within a transposon end sequence recognized by a
PT transposase useful for producing deletion derivatives of polypeptide.
XX
PS Claim 6; Page 31; Opp; English.
XX
CC The invention relates to a transposon nucleic acid comprising a
CC genetically engineered translation stop signal in the three reading
CC frames at least partly within a transposon end sequence recognized by a
CC transposase. The transposon is useful for producing deletion derivatives
CC of polypeptide coding nucleic acids. The method involves performing a
CC transposition reaction in the presence of a target nucleic acid
CC containing a polypeptide coding nucleic acid and in the presence of a
CC transposon containing a genetically engineered translation stop signal
CC sequence in three reading frames at least partly within a transposon end
CC sequence recognized by a transposase; and recovering a target nucleic
CC acid having the transposon incorporated in the protein coding nucleic
CC acid. The present sequence represents a Cat-Mu transposon modified end
CC fragment without 5' overhang
XX
SQ Sequence 50 BP; 21 A; 8 C; 11 G; 10 T; 0 U; 0 Other;

Query Match 92.6%; Score 50; DB 10; Length 50;
Best Local Similarity 100.0%; Pred. No. 1.5e-08;
Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 5 TGATTGATTGAACGAAACGCGAAGCGTTTCACGATAAATGCGAAAC 54
|||||
Db 1 TGATTGATTGAACGAAACGCGAAGCGTTTCACGATAAATGCGAAAC 50

RESULT 5
AAD21279
ID AAD21279 standard; DNA; 54 BP.
XX
AC AAD21279;
XX
DT 11-SEP-2003 (revised)
DT 28-JAN-2002 (first entry)
XX
DE Precut transposon end of Bacteriophage Mu non-transferred strand.
XX
KW Insertional mutation; synaptic complex; transposon; screening; ds.
XX
OS Enterobacteria phage Mu.
XX
PN US6294385-B1.
XX

PD ACF58170;
XX
DT 15-JAN-2004 (first entry)
XX
DE Cat-Mu transposon containing wild-type Mu ends.
XX
KW Transposon; genetic engineering; transposase; Cat-Mu; ds.
XX
OS Bacteriophage mu.
XX
FH Key Location/Qualifiers
FT misc_feature 1..4
FT /*tag= a
FT /note= "the 5' end of this strand overhangs the 3' end of
FT the complementary strand"
XX
PN WO2003087370-A1.
XX
PD 23-OCT-2003.
XX
PF 14-APR-2003; 2003WO-FI000285.
XX
PR 18-APR-2002; 2002FI-00000746.
XX
PA (FINN-) FINNZYMES OY.
XX
PI Savilahti H, Tieaho V;
XX

Query Match 79.3%; Score 42.8; DB 4; Length 54;
Best Local Similarity 87.0%; Pred. No. 6.4e-06;
Matches 47; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Oy 1 GATCTGATTGATTCAGCAACGAAACGCGAAGCGTTTCACGATAAATGCGAAAC 54
|||||
Db 1 GATCTGAAAGCGCGCAGCAACGCGAAGCGTTTCACGATAAATGCGAAAC 54

RESULT 6
ACF58170
ID ACF58170 standard; DNA; 54 BP.
XX
AC ACF58170;
XX
DT 15-JAN-2004 (first entry)
XX
DE Cat-Mu transposon containing wild-type Mu ends.
XX
KW Transposon; genetic engineering; transposase; Cat-Mu; ds.
XX
OS Bacteriophage mu.
XX
FH Key Location/Qualifiers
FT misc_feature 1..4
FT /*tag= a
FT /note= "the 5' end of this strand overhangs the 3' end of
FT the complementary strand"
XX
PN WO2003087370-A1.
XX
PD 23-OCT-2003.
XX
PF 14-APR-2003; 2003WO-FI000285.
XX
PR 18-APR-2002; 2002FI-00000746.
XX
PA (FINN-) FINNZYMES OY.
XX
PI Savilahti H, Tieaho V;
XX

```


AC ACC80742;

transposition of the Mu-end nucleic acid into the target, where transposition at the predominant site indicates the presence of mismatch at about that site. The methods are useful for typing a pathogenic microorganism strain, for screening an embryo for the presence of a mutation, for detecting the presence of a known mutation in a gene of interest, detecting the presence of a previously unidentified mutation in a gene of interest, and diagnosing the presence or absence of a tumour-promoting mutation. The kit is useful for detecting the presence of a mutation or polymorphism of interest in a nucleic acid molecule. The present sequence represents the Mu end DNA fragment construction oligonucleotide MM1141.

SQ Sequence 51 BP; 6 A; 13 C; 11 G; 21 T; 0 U; 0 Other;

Query Match 73.0%; Score 39.4; DB 13; Length 51;
Best Local Similarity 97.6%; Pred. No. 0.00011;
Matches 40; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 14 GAACGAAAACGCGAAGCGTTTCACGATAAATCGGAAAAC 54
Db 42 GCACGAAAACGCGAAGCGTTTCACGATAAATCGGAAAAC 2

RESULT 14

ADSI3820
ID ADSI3820 standard; DNA; 56 BP.

AC ADSI3820;

DT 16-DEC-2004 (first entry)

DE Mu end DNA fragment construction oligonucleotide MM1138.

KW ss: mismatch detection; transposition detection; pathogen typing;
KW embryo screening; mutation detection; Mu end DNA; MM1138.

OS Enterobacteria phage Mu.

OS Synthetic.

PN US2004191821-A1.

PD 30-SEP-2004.

PF 26-MAR-2004; 2004US-00809688.

PR 28-MAR-2003; 2003US-0457934P.

PA (USSA) US SEC OF ARMY.

PI Yanagihara K, Mizuuchi K;

DR WPI; 2004-689846/67.

XX Detecting a mismatch in a test double stranded nucleic acid target,
XX useful for typing a pathogenic microorganism strain, comprises detecting
XX transposition of the Mu-end nucleic acid into the target.

PS Example 1; SEQ ID NO 1; 24pp; English.

XX The invention relates to a method of detecting a mismatch in a test
XX double stranded nucleic acid target which comprises detecting
XX transposition of the Mu-end nucleic acid into the target, where
XX transposition at the predominant site indicates the presence of mismatch
XX at about that site. The methods are useful for typing a pathogenic
XX microorganism strain, for screening an embryo for the presence of
XX mutation, for detecting the presence of a known mutation in a gene of
XX interest, detecting the presence of a previously unidentified mutation in
XX a gene of interest, and diagnosing the presence or absence of a tumour-
XX promoting mutation. The kit is useful for detecting the presence of a
XX mutation or polymorphism of interest in a nucleic acid molecule. The
XX present sequence represents the Mu end DNA fragment construction
XX oligonucleotide MM1138.

SQ Sequence 56 BP; 22 A; 12 C; 15 G; 7 T; 0 U; 0 Other;

Query Match 73.0%; Score 39.4; DB 13; Length 56;
Best Local Similarity 97.6%; Pred. No. 0.00011;
Matches 40; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 14 GAACGAAAACGCGAAGCGTTTCACGATAAATCGGAAAAC 54
Db 15 GCACGAAAACGCGAAGCGTTTCACGATAAATCGGAAAAC 55

RESULT 15

AAZ28883/c
ID AAZ28883 standard; cDNA; 58 BP.

AC AAZ28883;

DT 15-FEB-2000 (first entry)

DE Mini-Mu transposable element deletion region delta-1.

KW Transposon Mu; transposable element; fusion protein; attachment site;
KW attL; attR; protein domain library; enzyme; accelerated evolution; ss.

OS Synthetic.

PN US5981177-A.

PD 09-NOV-1999.

PF 25-JAN-1995; 95US-00378548.

PR 25-JAN-1995; 95US-00378548.

PA (DEMI/) DEMIRJIAN D C.

PA (CASA/) CASADABAN M J.

PA (WEBE/) WEBER J M.

PA (GAIN/) GAINES G L.

PI Casadaban MJ, Demirjian DC, Weber JM, Gaines GL;

DR WPI; 1999-633307/54.

XX Generating fusion proteins using transposable elements, useful for
XX development of a protein domain library and in the construction of multi-
XX functional enzymes.

PS Example 2; Fig 2; 41pp; English.

XX The invention relates to a Mu-like transposable element (I) used for
XX generating functional fusion proteins after insertion into a target DNA.
XX The Mu-like element comprises: (a) a left transposable element attachment
XX site attL and a right transposable element attachment site attR, where
XX attR is no more than 62 nucleotides long; (b) a site for insertion of an
XX exogenous DNA sequence encoding for a protein domain located between attL
XX and attR; (c) after insertion of the transposable element into a target
XX DNA sequence, a fusion mRNA sequence is transcribed originating either
XX from the target DNA on either side of the transposable element or from
XX inside the transposable element and continuing through the attachment
XX site sequences and into the protein coding region, resulting in a single
XX fusion open reading frame (ORF). The constructs are useful in the
XX development of a protein domain library, in the construction of multi-
XX functional enzymes and in the accelerated evolution of new enzymatic
XX activities. The sequences AAZ28880-28884 represent deletion mini-Mu
XX elements of the invention (encoded ORF - AAY55901-Y55906)

SQ Sequence 58 BP; 8 A; 15 C; 13 G; 22 T; 0 U; 0 Other;

Query Match 73.0%; Score 39.4; DB 2; Length 58;
Best Local Similarity 97.6%; Pred. No. 0.00011;
Matches 40; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 14 GAACGAAAACGCGAAGCGTTTCACGATAAATCGGAAAAC 54

Db 49 GCACGAAAAACCGCGAAGCGTTTCACGATAAATCGGAAAC 9

Search completed: January 17, 2006, 19:35:05
Job time : 35.5794 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2006 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 17, 2006, 18:09:28 ; Search time 228.555 Seconds
(without alignments)
11054.239 Million cell updates/sec

Title: US-10-511-327-1
Perfect score: 54
Sequence: 1 gatctgattgtaacgaa.....ttcacgataaatgcgaaaac 54

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues
Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:
1: gb_est1:
2: gb_est2:
3: gb_est3:
4: gb_hic:
5: gb_est4:
6: gb_est5:
7: gb_est6:
8: gb_est7:
9: gb_ges1:
10: gb_ges2:
11: gb_ges3:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|--------------------|
| 1 | 40.4 | 74.8 | 370 | 10 | CZ569838 OB_Ba000 |
| 2 | 39.4 | 73.0 | 226 | 9 | CC742071 ZMWB0011 |
| 3 | 39.4 | 73.0 | 589 | 10 | CL858345 OR_CBA008 |
| 4 | 36.2 | 67.0 | 543 | 10 | CZ821413 OC_Ba019 |
| 5 | 36.2 | 67.0 | 688 | 10 | CZ820900 OC_Ba019 |
| 6 | 30.6 | 56.7 | 247 | 10 | CZ821171 OC_Ba019 |
| 7 | 27.4 | 50.7 | 582 | 3 | BJ016030 BJ016030 |
| 8 | 26.6 | 49.3 | 757 | 3 | BI953957 HVSMEM001 |
| 9 | 26.6 | 49.3 | 802 | 10 | D0032104 13515 Tom |
| 10 | 26.4 | 48.9 | 274 | 8 | T31184 EST28439 Hu |
| 11 | 26.4 | 48.9 | 771 | 3 | BI953902 HVSMEM001 |
| 12 | 26.2 | 48.5 | 215 | 3 | BI953982 HVSMEM001 |
| 13 | 26.2 | 48.5 | 771 | 3 | BI953934 HVSMEM001 |
| 14 | 26 | 48.1 | 384 | 6 | CD113955 MEL-0032P |
| 15 | 26 | 48.1 | 510 | 9 | BH821089 BACPP15-D |
| 16 | 26 | 48.1 | 737 | 3 | BI954000 HVSMEM001 |
| 17 | 25.8 | 47.8 | 248 | 8 | F05165 HSC02H021 n |
| 18 | 25.8 | 47.8 | 304 | 1 | AA481614 aa35g09.r |
| 19 | 25.8 | 47.8 | 403 | 1 | AA262843 zs24d05.r |
| 20 | 25.8 | 47.8 | 535 | 5 | C14413 C14413 Clon |
| 21 | 25.8 | 47.8 | 561 | 1 | AW954984 EST367054 |
| 22 | 25.8 | 47.8 | 634 | 1 | AW955329 EST367399 |

| | | | | | | |
|------|------|------|-----|----|----------|--------------------|
| 23 | 25.8 | 47.8 | 699 | 9 | A2571139 | A2571139 283PvD05 |
| C 24 | 25.8 | 47.8 | 731 | 1 | AL041260 | AL041260 DKF2p434L |
| C 25 | 25.8 | 47.8 | 856 | 3 | BI772228 | BI772228 603056184 |
| C 26 | 25.8 | 47.8 | 994 | 3 | BM804853 | BM804853 AGENCOURT |
| C 27 | 25.6 | 47.4 | 196 | 3 | BI953931 | BI953931 HVSMEM001 |
| C 28 | 25.6 | 47.4 | 705 | 3 | BO149276 | BO149276 NF088G08F |
| C 29 | 25.6 | 47.4 | 752 | 3 | BI953878 | BI953878 HVSMEM001 |
| C 30 | 25.6 | 47.4 | 973 | 10 | D0005263 | D0005263 300893 To |
| C 31 | 25.4 | 47.0 | 484 | 9 | A2152875 | A2152875 SP_0046_B |
| C 32 | 25.4 | 47.0 | 614 | 7 | CJ393910 | CJ393910 CJ393910 |
| C 33 | 25.4 | 47.0 | 638 | 7 | CJ417250 | CJ417250 CJ417250 |
| C 34 | 25.4 | 47.0 | 749 | 3 | BI953959 | BI953959 HVSMEM001 |
| C 35 | 25.4 | 47.0 | 887 | 7 | CN586362 | CN586362 USDA-FP_1 |
| C 36 | 25.2 | 46.7 | 297 | 6 | CF504687 | CF504687 USDA-FP_1 |
| C 37 | 25.2 | 46.7 | 491 | 8 | DN797614 | DN797614 USDA-FP_1 |
| C 38 | 25.2 | 46.7 | 518 | 9 | BH684386 | BH684386 BOHP87TF |
| C 39 | 25.2 | 46.7 | 671 | 9 | BZ833089 | BZ833089 CH240_222 |
| C 40 | 25.2 | 46.7 | 700 | 8 | DR400612 | DR400612 TKN060506 |
| C 41 | 25.2 | 46.7 | 703 | 5 | BW085322 | BW085322 BW085322 |
| C 42 | 25.2 | 46.7 | 706 | 3 | BM617600 | BM617600 170006871 |
| C 43 | 25.2 | 46.7 | 741 | 3 | BI953998 | BI953998 HVSMEM001 |
| C 44 | 25.2 | 46.7 | 748 | 3 | BI954023 | BI954023 HVSMEM001 |
| C 45 | 25.2 | 46.7 | 776 | 6 | CB292852 | CB292852 UCRC501_0 |

ALIGNMENTS

RESULT 1
CZ569838
LOCUS
DEFINITION OB_Ba0003F02.r OB_Ba Oryza brachyantha genomic clone
ACCESSION OB_Ba0003F02 3', Genomic survey sequence.
VERSION CZ569838.1 GI:68013639
KEYWORDS GSS.
SOURCE Oryza brachyantha
ORGANISM Oryza brachyantha
REFERENCE
AUTHORS Kim,H., Yu,Y., Stum,D., Yost,D., Rao,K., Luo,M., Jetty,R.,
Kudrna,D., Muller,C., Hatfield,J., Soderlund,C. and Wing,R.
TITLE OMAP Project
JOURNAL Unpublished (2004)
COMMENT Contact: Rod A. Wing
Arizona Genomics Institute
University of Arizona
Forbes Building Room 303, Tucson, AZ 85721-0036, USA
Tel: 520 626 9595
Fax: 520 621 1259
Email: rwing@genome.arizona.edu
PCR Primers
FORWARD: TAA TAC GAC TCA CTA TAG GG
BACKWARD: CAC TCA TTA GGC ACC CCA
Insert Length: 145 Std Error: 0.00
Plate: 0003 row: F column: 02
Seq primer: CAC TCA TTA GGC ACC CCA
Class: BAC ends.
Location/Qualifiers
1. 370
/organism="Oryza brachyantha"
/mol_type="genomic DNA"
/db_xref="taxon:4533"
/clone="OB_Ba0003F02"
/tissue_type="leaves"
/dev_stage="mature"
/lab_host="DH10B"
/clone_lib="OB_Ba"
/note="Vector: pGIBAC1; Site_1: HindIII; Site_2: HindIII"

```
Query Match          74.8%; Score 40.4; DB 10; Length 370;
Best Local Similarity 97.6%; Pred. No. 0.00042;
Matches 41; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 13 TGAACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAAC 54
    |||||||
Db 85 TGCACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAAC 126

RESULT 2
CC742071
LOCUS CC742071.1 ZMWBb0115J07.f ZMWBb Zea mays genomic clone ZMWBb0115J07 5',
DEFINITION genomic survey sequence.
ACCESSION CC742071.1 GI:32194524
VERSION CC742071.1
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 226)
AUTHORS Yu, Y., Kim, H. R., Hatfield, J., Soderlund, C., Bharti, A. K., Messing, J.,
and Wing, R.
TITLE Sequencing of the maize genome
JOURNAL Unpublished (2003)
COMMENT Contact: Rod Wing
Arizona Genomics Institute
University of Arizona
Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ
85721-0088, USA
Tel: 520 626 3967
Fax: 520 621 9288
Email: rwing@genome.arizona.edu
PCR Primers
FORWARD: T7
BACKWARD: M13r
Plate: 0115 row: J column: 07
Seq primer: T7
Class: BAC ends.
FEATURES
    source
        Location/Qualifiers
            1..226
                /organism="Zea mays"
                /mol_type="genomic DNA"
                /cultivar="B73"
                /db_xref="taxon:4577"
                /clone="ZMWBb0115J07"
                /lab_host="DH10B"
                /clone_lib="ZMWBb"
                /note="Vector: pBelOBAC11; Site_1: HindIII; Site_2:
                HindIII; Zea mays L. ssp. mays"

ORIGIN
Query Match          73.0%; Score 39.4; DB 9; Length 226;
Best Local Similarity 97.6%; Pred. No. 0.00095;
Matches 40; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 14 GAACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAAC 54
    |||||||
Db 182 GCACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAAC 222

RESULT 3
CL858345
LOCUS CL858345.1 OR_CBa Oryza rufipogon genomic clone OR_CBa0089D12
DEFINITION 5', genomic survey sequence.
ACCESSION CL858345
VERSION CL858345.1 GI:51269584
KEYWORDS GSS.
SOURCE Oryza rufipogon
ORGANISM Oryza rufipogon
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.
REFERENCE 1 (bases 1 to 543)
AUTHORS Kim, H., Collura, K., Wissotski, M., Byrne, M., Stum, D., Smart, D.,
Rao, K., Luo, M., Jetty, R., Kudrna, D., Muller, C., Soderlund, C. and
Wing, R.
TITLE Oryza Map Alignment Project - Arizona Genomics Institute
JOURNAL Unpublished (2005)
COMMENT Contact: Rod A. Wing
Arizona Genomics Institute
University of Arizona
Forbes Building Room 303, Tucson, AZ 85721-0036, USA
Tel: 520 626 9595
Fax: 520 621 1259
Email: rwing@genome.arizona.edu
PCR Primers
FORWARD: TAA TAC GAC TCA CTA TAG GG
BACKWARD: CAC TCA TTA GGC ACC CCA
Plate: 0089 row: D column: 12
Seq primer: TAA TAC GAC TCA CTA TAG GG
Class: BAC ends.
FEATURES
    source
        Location/Qualifiers
            1..589
                /organism="Oryza rufipogon"
                /mol_type="genomic DNA"
                /db_xref="taxon:4529"
                /clone="OR_Cha0089D12"
                /tissue_type="young leaves"
                /dev_stage="2 week old seedlings"
                /lab_host="DH10B T1 phage resistant"
                /clone_lib="OR_Cha"
                /note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII;
                drk treated 36 hrs before harvest"

ORIGIN
Query Match          73.0%; Score 39.4; DB 10; Length 589;
Best Local Similarity 97.6%; Pred. No. 0.00098;
Matches 40; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 14 GAACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAAC 54
    |||||||
Db 495 GCACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAAC 535

RESULT 4
CZ821413
LOCUS CZ821413.1 OC_Ba Oryza coarctata genomic clone OC_Ba0199P13
DEFINITION 3', genomic survey sequence.
ACCESSION CZ821413
VERSION CZ821413.1 GI:71261266
KEYWORDS GSS.
SOURCE Oryza coarctata (Porteresia coarctata)
ORGANISM Oryza coarctata
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.
REFERENCE 1 (bases 1 to 543)
AUTHORS Kim, H., Collura, K., Wissotski, M., Byrne, M., Stum, D., Smart, D.,
Rao, K., Luo, M., Jetty, R., Kudrna, D., Muller, C., Soderlund, C. and
Wing, R.
TITLE Oryza Map Alignment Project - Arizona Genomics Institute
JOURNAL Unpublished (2005)
COMMENT Contact: Rod A. Wing
Arizona Genomics Institute
University of Arizona
Forbes Building Room 303, Tucson, AZ 85721-0036, USA
Tel: 520 626 9595
Fax: 520 621 1259
Email: rwing@genome.arizona.edu
PCR Primers
FORWARD: TAA TAC GAC TCA CTA TAG GG
```

```
BACKWARD: CAC TCA TTA GGC ACC CCA
Plate: 0199 row: P column: 13
Seq primer: CAC TCA TTA GGC ACC CCA
Class: BAC ends.

FEATURES
    source
        1..543
            /organism="Oryza coarctata"
            /mol_type="genomic DNA"
            /db_xref="taxon:77588"
            /clone="OC_Ba0199p13"
            /tissue_type="leaves"
            /dev_stage="mature"
            /lab_host="DH10B"
            /clone_lib="OC_Ba"
            /note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII"

ORIGIN
Query Match      67.0%; Score 36.2; DB 10; Length 543;
Best Local Similarity 92.7%; Pred. No. 0.014;
Matches 38; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 14 GAACGAAACCGGAAAGCGTTTCACGATAAATGCGAAAC 54
    |||||||
Db 473 GCACGAAATGCGAAAGCGTTTCAGTACATGCGAAAC 513

RESULT 5
CZ820900      688 bp DNA linear GSS 26-JUL-2005
LOCUS OC_Ba0199D21.r OC_Ba Oryza coarctata genomic clone OC_Ba0199D21
DEFINITION 3', genomic survey sequence.
ACCESSION CZ820900
VERSION CZ820900.1 GI:71260753
KEYWORDS GSS.
SOURCE Oryza coarctata (Porteresia coarctata)
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.
REFERENCE 1 (bases 1 to 688)
AUTHORS Kim,H., Collura,K., Wisotski,M., Byrne,M., Stum,D., Smart,D.,
Rao,K., Luo,M., Jetty,R., Kudrna,D., Muller,C., Soderlund,C. and
Wing,R.
TITLE OMAP (Oryza Map Alignment Project)- Arizona Genomics Institute
JOURNAL Unpublished (2005)
COMMENT Contact: Rod A. Wing
Arizona Genomics Institute
University of Arizona
Forbes Building Room 303, Tucson, AZ 85721-0036, USA
Tel: 520 626 9595
Fax: 520 621 1259
Email: rwing@genome.arizona.edu
PCR Primers
FORWARD: TAA TAC GAC TCA CTA TAG GG
BACKWARD: CAC TCA TTA GGC ACC CCA
Plate: 0199 row: D column: 21
Seq primer: CAC TCA TTA GGC ACC CCA
Class: BAC ends.
FEATURES
    source
        1..247
            /organism="Oryza coarctata"
            /mol_type="genomic DNA"
            /db_xref="taxon:77588"
            /clone="OC_Ba0199J21"
            /tissue_type="leaves"
            /dev_stage="mature"
            /lab_host="DH10B"
            /clone_lib="OC_Ba"
            /note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII"

ORIGIN
Query Match      56.7%; Score 30.6; DB 10; Length 247;
Best Local Similarity 73.6%; Pred. No. 1.5;
Matches 39; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

Oy 2 ATCTGATTGATTGAACGAAACCGGAAAGCGTTTCACGATAAATGCGAAAC 54
    |||||
Db 132 ATATGAAGCGGTGCATCAAAAATCCGAAGCGTTTCATGATCATGTGAAAC 184

RESULT 7
BJ016030      582 bp mRNA linear EST 05-DEC-2001
LOCUS BJ016030
DEFINITION BJ016030 MF01SSA cDNA Oryzias latipes cDNA clone MF01SSA008A02 3',
mRNA sequence.
ACCESSION BJ016030
VERSION BJ016030.1 GI:17375178
KEYWORDS EST.
SOURCE Oryzias latipes (Japanese medaka)
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
```

Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
 Belontiiformes; Adrianichthyidae; Oryziinae; Oryzias.
 1 (bases 1 to 582)
REFERENCE
AUTHORS Kohara,Y., Shin-I,T., Kimura,T., Narita,T., Jindo,T. and Takeda,H.
TITLE Medaka EST Project in Takeda's lab
JOURNAL Unpublished (2001)
COMMENT Contact: Tadao Shin-i
 Center For Genetic Resource Information
 National Institute of Genetics
 1111 Yata, Mishima, Shizuoka 411-8540, Japan
 Tel: 81-559-81-6856
 Fax: 81-559-81-6855
 Email: tshini@genes.nig.ac.jp.

FEATURES
 source
 1..582
 Location/Qualifiers
 1..582
 /organism="Oryzias latipes"
 /mol_type="mRNA"
 /strain="Hd-rR"
 /db_xref="taxon:8090"
 /clone="MF01SSA008A02"
 /sex="mixture of female and male"
 /tissue_type="whole embryo"
 /dev_stage="segmentation stage 20 - 25"
 /clone_lib="MF01SSA cDNA"

ORIGIN

Query Match 50.7%; Score 27.4; DB 3; Length 582;
 Best Local Similarity 69.8%; Pred. No. 23;
 Matches 37; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

Qy 1 GATCGATTGATTGAACGAAACCGAAAGCGTTTCACGATAAATGCGAAA 53
 Db 206 GTTGTCATTATTAAACGTGAAACCAAAAGCATTTTCAGATAATACTACA 258

RESULT 8
 BI953957/c
 LOCUS HVSMEM0015112f 757 bp mRNA linear EST 19-OCT-2001
 DEFINITION HVCDNA0014 (Blumeria infected) Hordeum vulgare subsp. vulgare cDNA
 clone HVSMEM0015112f, mRNA sequence.

ACCESSION BI953957
 VERSION BI953957.1 GI:16299057
 KEYWORDS EST.

SOURCE Hordeum vulgare subsp. vulgare
 ORGANISM Hordeum vulgare subsp. vulgare
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Poideae; Triticeae; Hordeum.

REFERENCE 1 (bases 1 to 757)
 AUTHORS Wing,R., Close,T.J., Kleinhofs,A., Wise,R., Chin,A., Begum,D.,
 Frisch,D., Atkins,M., Yu,Y., Henry,D., Palmer,M., Rambo,T.,
 Simmons,J., Oates,R. and Main,D.
 TITLE Development of a genetically and physically anchored EST resource
 for barley genomics: Blumeria infected Morex (compatible) seedling
 cDNA library

JOURNAL Unpublished (2001)
 COMMENT Contact: Wing RA
 Clemson University Genomics Institute
 Clemson University
 100 Jordan Hall, Clemson, SC 29634, USA

Tel: 864 656 7288
 Fax: 864 656 4293
 Email: rwing@clemson.edu
 Total hq bases = 590
 Seq primer: AATTAACTCTACTAAAGGG
 High quality sequence start: 4
 High quality sequence stop: 756.

FEATURES
 source
 1..757
 /organism="Hordeum vulgare subsp. vulgare"
 /mol_type="mRNA"
 /cultivar="Morex"

/sub_species="vulgare"
 /db_xref="taxon:112509"
 /clone="HVSMEM0015112f"
 /tissue_type="green seedling leaf"
 /lab_host="TJC121"
 /clone_lib="Hordeum vulgare green seedling EST library
 HVCDNA0014 (Blumeria infected)"
 /note="Vector: pBluescript SK(-); Site 1: EcoRI; Site 2:
 XhoI; Morex (mla) plants were greenhouse grown in the R
 Wise lab at Iowa State University, Ames, IA; 7 day old
 green seedlings were infected with isolate 5874 of
 Blumeria graminis f. sp. hordei, and leaves were harvested
 24, 48 and 72 hr post-inoculation and snap frozen (Wise).
 In the TJ Close lab at the University of California,
 Riverside, total RNA was prepared from each sample pool,
 equal quantities of all three RNA pools were combined,
 poly(A) RNA was purified from the mixture, one primary
 unamplified cDNA library was made, and 1 million pfu were
 in vivo excised to give pBluescript SK(-) cDNA phagemids
 (Chin). Phagemids were plated and picked at the Clemson
 University Genomics Institute (CUGI) (Begum, Palmer,
 Frisch, Atkins and Wing). Plasmid DNA preparations, DNA
 sequencing and sequence analysis were performed at CUGI
 (Wing, Yu, Frisch, Henry, Simmons, Oates, Rambo, Main).
 The sequence has been trimmed to remove vector sequence
 and contains a minimum of 100 bases of phred value 20 or
 above. For more details on library preparation and
 sequence analysis see
 http://www.genome.clemson.edu/projects/barley. To order
 this clone see http://www.genome.clemson.edu/orders/Also
 see Close TJ, Wing R, Kleinhofs A, Wise R (2001)
 Genetically and physically anchored EST resources for
 barley genomics. Barley Genetics Newsletter 31:29-30.
 (http://wheat.pw.usda.gov/gppages/bgn/31/cover.html)"

ORIGIN

Query Match 49.3%; Score 26.6; DB 3; Length 757;
 Best Local Similarity 81.1%; Pred. No. 46;
 Matches 43; Conservative 0; Mismatches 9; Indels 1; Gaps 1;
 Qy 2 ATCTGATTGATTGAACGAAACCGAAAGCGTTTCACGATAAATGCGAAAAC 54
 Db 52 AGCTGAAGCGCGCCACGAAAAACGCGAAAGCTTTTCACG-TAAATGCGAAAAC 1

RESULT 9

DU032104/c

LOCUS

DEFINITION

13515 Tomato HindIII BAC Library Lycopersicon esculentum genomic

clone LE_HBa0169A20 3, genomic survey sequence.

ACCESSION DU032104.1 GI:72445405

VERSION

KEYWORDS

SOURCE

ORGANISM

Lycopersicon esculentum (Solanum lycopersicum)

Lycopersicon esculentum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicotyledons;

asterids; lamids; Solanales; Solanaceae; Solanum; Lycopersicon.

REFERENCE 1 (bases 1 to 802)

AUTHORS Mueller,L.A., Buels,R.M., Wang,Y., Tanksley,S.D., Giovannoni,J.J.,

Van Eck,J. and Stack,S.

TITLE BAC end sequencing from three Solanum lycopersicon libraries

JOURNAL Unpublished (2005)

COMMENT Other_GSSs: 13518

Contact: Lukas Mueller

Tanksley Lab, Dept. of Plant Breeding

Cornell University

251 Emerson Hall, Ithaca, NY 14853, USA

Tel: 607-255-6557

Fax: 607-255-6683

Email: sgn-feedback@sgn.cornell.edu

Plate: 169 row: A column: 20

Seq primer: SP6


```

Class: BAC ends
High quality sequence start: 36
High quality sequence stop: 428.

FEATURES
    source
        1..802
            /organism="lycopersicon esculentum"
            /mol_type="genomic DNA"
            /cultivar="Heinz 1706"
            /db_xref="taxon:4081"
            /clone="LE HBa0169A20"
            /lab_host="E. coli"
            /clone_lib="Tomato HindIII BAC Library"
            /note="Vector: pBelOBAC1; Site_1: HindIII"

ORIGIN
Query Match      49.3%; Score 26.6; DB 10; Length 802;
Best Local Similarity 71.4%; Pred. No. 46;
Matches 35; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY   6 GATTGATTAAGCGAAACGGGAAGCGTTTCACGATAAATGCAGAAAAC 54
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB   693 GTGAATTAAAAAGAAAGACGAGATTTCCTAAATATGAGAAAAC 645

RESULT 10
T31184/c
LOCUS       T31184.1 GI:613282
DEFINITION Homo sapiens (human)
ACCESSION  T31184
VERSION     T31184.1
KEYWORDS    EST.
SOURCE      Homo sapiens
ORGANISM    Homo sapiens
REFERENCE   1 (bases 1 to 274)
AUTHORS     Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A.,
Bult,C.J., Lee,N., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D.,
White,O., Sutton,G., Blake,J.A., Brandon,R.C., Chiu,M.-W.,
Clayton,R.A., Cline,R.T., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,
Fitzgerald,L.M., FitzHugh,W.M., Fritchman,J.L., Geoghegan,N.S.M.,
Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., HinkleJr,P.S.,
Kelley,J.M., Klinek,K., Kelley,J.C., Liu,L.-I., Marmaros,S.M.,
Merrick,J.M., Moreno-Palauques,R.F., McDonald,L.A., Nguyen,D.T.,
Pellegriano,S.M., Phillips,C.A., Ryder,S.E., Scott,J.L.,
Saudek,D.M., Shirley,R., Smali,K.V., Spriggs,T.A., Utterback,T.R.,
Weidman,J.F., Li,Y., Bednarek,D.P., Cao,L., Cepeda,M.A.,
Coleman,T.A., Collins,E.-J., Dimke,D., Feng,P., Ferrie,A.,
Fischer,C., Hastings,G.A., He,W.-W., Hu,J.-S., Greene,J.M.,
Gruber,J., Hudson,P., Kim,A., Kozak,D.L., Kunsch,C., Ji,H., Di,H.,
Meisner,P.S., Olsen,H., Raymond,L.J., Wei,Y.-F., Wing,J., Xu,C.,
Yu,G.-L., Ruben,S.M., Dillon,P.J., Fannon,M.R., Rosen,C.A.,
Haseltine,W.A., Fields,C., Fraser,C.M. and Venter,J.C.
Initial Assessment of Human Gene Diversity and Expression Patterns
Based Upon 83 Million Basepairs of cDNA Sequence
Nature 377, 3-174 (1995)
7566098
Contact: Venter, JC
The Institute for Genomic Research
932 Clopper Rd, Gaithersburg, MD 20878
Tel: 3018699056
Fax: 3018699423
Email: tdbinfo@tdb.tigr.org
For clone availability, additional sequence and expression
information related to this EST, please contact the TIGR Database
(tdbinfo@tdb.tigr.org)
Seq primer: M13 Reverse.
Location/Qualifiers
1..274
/organism="Homo sapiens"
/mol_type="mRNA"

FEATURES
    source

```

University Genomics Institute (CUGI) (Begum, Palmer, Frisch, Atkins and Wing). Plasmid DNA preparations, DNA sequencing and sequence analysis were performed at CUGI (Wing, Yu, Frisch, Henry, Simmons, Oates, Rambo, Main). The sequence has been trimmed to remove vector sequence and contains a minimum of 100 bases of phred value 20 or above. For more details on library preparation and sequence analysis see <http://www.genome.clemson.edu/projects/barley>. To order this clone see <http://www.genome.clemson.edu/orders> Also see Close TJ, Wing R, Kleinohs A, Wise R (2001) Genetically and physically anchored EST resources for barley genomics. Barley Genetics Newsletter 31:29-30. (<http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html>)"

ORIGIN

Query Match 48.9%; Score 26.4; DB 3; Length 771;
Best Local Similarity 75.0%; Pred. No. 54;
Matches 33; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Qy 2 ATCTGATTGTTGAACGAAAAACGCGAAGCGTTTCACGATAAA 45
|||
Db 44 ATGTGACGGCGCAGAAAAACGCGAAGCGTTTCACGATAAA 1

RESULT 12

BI953982/c
LOCUS BI953982 215 bp mRNA linear EST 19-OCT-2001
DEFINITION HVSMEM0015K12f Hordeum vulgare green seedling EST library
clone HVSMEM0015K12f, mRNA sequence.

ACCESSION

VERSION BI953982.1 GI:16299103

KEYWORDS

SOURCE EST.

ORGANISM

Hordeum vulgare subsp. vulgare
Hordeum vulgare subsp. vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Pooideae; Triticeae; Hordeum.
1 (bases 1 to 215)
Wing, R., Close, T.J., Kleinohs, A., Wise, R., Chin, A., Begum, D.,
Frisch, D., Atkins, M., Yu, Y., Henry, D., Palmer, M., Rambo, T.,
Simmons, J., Oates, R. and Main, D.
Development of a genetically and physically anchored EST resource
for barley genomics: Blumeria infected Morex (compatible) seedling
cDNA library

JOURNAL

COMMENT

Unpublished (2001)
Contact: Wing RA
Clemson University
Clemson University Genomics Institute
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Total hg bases = 174
Seq primer: AATTAACCTCTACTAAAGGG
High quality sequence start: 4
High quality sequence stop: 216.
Location/Qualifiers

FEATURES

source

1..215
/organism="Hordeum vulgare subsp. vulgare"
/mol_type="mRNA"
/cultivar="Morex"
/sub_species="vulgare"
/db_xref="taxon:112509"
/clone="HVSMEM0015K12f"
/tissue_type="green seedling leaf"
/lab_host="TJG121"
/clone_lib="Hordeum vulgare green seedling EST library
HVCDNA0014 (Blumeria infected)"
/note="Vector: pBluescript SK(-); Site 1: EcoRI; Site 2:
XhoI; Morex (mla) plants were greenhouse grown in the R
Wise lab at Iowa State University, Ames, IA; 7 day old

ORIGIN

Query Match 48.5%; Score 26.2; DB 3; Length 215;
Best Local Similarity 79.5%; Pred. No. 62;
Matches 31; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 14 GAACGAAAAACGCGAAGCGTTTCACGATAAATCGGAAA 52
|||
Db 40 GCACGAAAAACGCGAAGCGTTTCACGATAAAGCGAAAA 2

RESULT 13

BI953934/c

LOCUS

DEFINITION BI953934 771 bp mRNA linear EST 19-OCT-2001
HVSMEM0015G22f Hordeum vulgare green seedling EST library
clone HVSMEM0015G22f, mRNA sequence.

ACCESSION BI953934.1 GI:16299017

VERSION

KEYWORDS

SOURCE

ORGANISM

Hordeum vulgare subsp. vulgare
Hordeum vulgare subsp. vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Pooideae; Triticeae; Hordeum.

REFERENCE

AUTHORS

1 (bases 1 to 771)

Wing, R., Close, T.J., Kleinohs, A., Wise, R., Chin, A., Begum, D.,

Frisch, D., Atkins, M., Yu, Y., Henry, D., Palmer, M., Rambo, T.,

Simmons, J., Oates, R. and Main, D.

Development of a genetically and physically anchored EST resource

for barley genomics: Blumeria infected Morex (compatible) seedling

cDNA library

Unpublished (2001)

Contact: Wing RA

Clemson University

Clemson University Genomics Institute

100 Jordan Hall, Clemson, SC 29634, USA

Tel: 864 656 7288

Fax: 864 656 4293

Email: rwing@clemson.edu

Total hg bases = 471

Seq primer: AATTAACCTCTACTAAAGGG

High quality sequence start: 25

High quality sequence stop: 736.

Location/Qualifiers

1..771

/organism="Hordeum vulgare subsp. vulgare"

/mol_type="mRNA"

/cultivar="Morex"

green seedlings were infected with isolate 5874 of
Blumeria graminis f. sp. hordei, and leaves were harvested
24, 48 and 72 hr post-inoculation and snap frozen (Wise).
In the TJ Close lab at the University of California,
Riverside, total RNA was prepared from each sample pool,
equal quantities of all three RNA pools were combined,
poly(A) RNA was purified from the mixture, one primary
unamplified cDNA library was made, and 1 million pfu were
in vivo excised to give pBluescript SK(-) cDNA phagemids
(Chin). Phagemids were plated and picked at the Clemson
University Genomics Institute (CUGI) (Begum, Palmer,
Frisch, Atkins and Wing). Plasmid DNA preparations, DNA
sequencing and sequence analysis were performed at CUGI
(Wing, Yu, Frisch, Henry, Simmons, Oates, Rambo, Main).
The sequence has been trimmed to remove vector sequence
and contains a minimum of 100 bases of phred value 20 or
above. For more details on library preparation and
sequence analysis see

<http://www.genome.clemson.edu/projects/barley>. To order
this clone see <http://www.genome.clemson.edu/orders> Also
see Close TJ, Wing R, Kleinohs A, Wise R (2001)
Genetically and physically anchored EST resources for
barley genomics. Barley Genetics Newsletter 31:29-30.
(<http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html>)"

/sub_species="vulgare"
 /db_xref="taxon:112509"
 /clone="HVS MEM0015G22f"
 /tissue_type="green seedling leaf"
 /lab_host="TJC121"
 /clone_lib="Hordeum vulgare green seedling EST library
 HVC DNA0014 (Blumeria infected)"
 /note="Vector: pBluescript SK(-); Site 1: EcoRI; Site 2:
 XhoI; Morex (mla) plants were greenhouse grown in the R
 wise lab at Iowa State University, Ames, IA; 7 day old
 green seedlings were infected with isolate 5974 of
 Blumeria graminis f. sp. hordei, and leaves were harvested
 24, 48 and 72 hr post-inoculation and snap frozen (Wise).
 In the TJ Close lab at the University of California,
 Riverside, total RNA was prepared from each sample pool,
 equal quantities of all three RNA pools were combined,
 poly(A) RNA was purified from the mixture, one primary
 unamplified cDNA library was made, and 1 million pfu were
 in vivo excised to give phagescript SK(-) cDNA phagemids
 (Chin). Phagemids were plated and picked at the Clemson
 University Genomics Institute (CUGI) (Begum, Palmer,
 Frisch, Atkins and Wing). Plasmid DNA preparations, DNA
 sequencing and sequence analysis were performed at CUGI
 (Wing, Yu, Frisch, Henry, Simmons, Oates, Rambo, Main).
 The sequence has been trimmed to remove vector sequence
 and contains a minimum of 100 bases of phred value 20 or
 above. For more details on library preparation and
 sequence analysis see
<http://www.genome.clemson.edu/projects/barley>. To order
 this clone see <http://www.genome.clemson.edu/orders> Also
 see Close TJ, Wing R, Kleinhofs A, Wise R (2001)
 Genetically and physically anchored EST resources for
 barley genomics. Barley Genetics Newsletter 31:29-30.
 (<http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html>)"

ORIGIN

Query Match 48.5%; Score 26.2; DB 3; Length 771;
 Best Local Similarity 79.5%; Pred. No. 64;
 Matches 31; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 14 GAACGAAAACGCGAAGCGTTTCAGTAAATGCGAA 52
 |||||
 DB 40 GCACGAAAACGCGAAGCGTTTCAGTAAATGCGAA 52
 |||||

RESULT 14

CD113955/c 384 bp mRNA linear EST 14-SEP-2003
 LOCUS ME1-0032P-A101-F03-U-B ME1-0032 Schistosoma mansoni cDNA clone
 DEFINITION ME1-0032P-A101-F03.B, mRNA sequence.
 ACCESSION CD113955
 VERSION CD113955.1 GI:34652145
 KEYWORDS EST
 SOURCE Schistosoma mansoni
 ORGANISM Schistosoma mansoni
 Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;
 Strigideida; Schistosomatoidea; Schistosomatidae; Schistosoma.
 1 (bases 1 to 384)
 AUTHORS Verjovski-Almeida, S., DeMarco, R., Martins, E.A.L., Guimaraes, P.E.M.,
 Ojopi, E.P.B., Paquola, A.C.M., Piazza, J.P., Nishiyama, M.Y. Jr.,
 Kitajima, J.P., Adamson, R.E., Ashton, P.D., Bonaldo, M.F.,
 Coulson, P.S., Dillon, G.P., Farias, L.P., Gregorio, S.P., Ho, P.L.,
 Leite, R.A., Malaquias, L.C.C., Marques, R.C.P., Miyasato, P.A.,
 Nascimento, A.L.T.O., Ohlweiler, F.P., Reis, E.M., Ribeiro, M.A.,
 Sa, R.G., Stukart, G.C., Soares, M.B., Gargioni, C., Kawano, T.,
 Rodrigues, V., Madeira, A.M.B.N., Wilson, R.A., Menck, C.F.M.,
 Setubal, J.C., Leite, L.C.C. and Dias-Neto, E.
 Transcriptome analysis of the acelomate human parasite Schistosoma
 mansoni

TITLE

JOURNAL Nat. Genet. 35 (2), 148-157 (2003)
 PUBMED 12973350
 COMMENT Contact: Dr. Sergio Verjovski-Almeida
 Departamento de Bioquímica

Instituto de Química - Universidade de Sao Paulo
 Av. Prof. Lineu Prestes 748 sala 1200, 05508-900 Sao Paulo - SP,
 Brasil

Tel: +55-11-3091-2173

Fax: +55-11-3091-2186

Email: verjovski@usp.br

This sequence was derived from the FAPESP Schistosoma mansoni EST
 Genome Project. All sequences in the project were assembled and
 annotated. This entry and all the assembled sequences can be seen
 in the following URL <http://bioinfo.iq.usp.br/schisto/>
 Plate: ME1-0032P-A101 row: 3 column: F.

FEATURES

source
 1..384
 Location/Qualifiers
 /organism="Schistosoma mansoni"
 /mol_type="mRNA"
 /db_xref="taxon:6183"
 /clone="ME1-0032P-A101-F03.B"
 /sex="mixed pool"
 /dev_stage="egg"
 /lab_host="Mus musculus"
 /clone_lib="ME1-0032"
 /note="Vector: pGEM T-easy"

ORIGIN

Query Match 48.1%; Score 26; DB 6; Length 384;
 Best Local Similarity 70.0%; Pred. No. 74;
 Matches 35; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 2 ATCTGATTGATGACGAAACGCGAAGCGTTTCAGTAAATGCGAA 51
 |||||

DB 50 ATCTCATCAATGAACGAAATGAAATGCGTTTAAAGACTGATGCTAA 1
 |||||

RESULT 15

BH821089 510 bp DNA linear GSS 20-MAY-2002
 LOCUS BH821089
 DEFINITION BH821089.1
 ACCESSION BH821089
 VERSION BH821089.1 GI:20996539
 KEYWORDS GSS.
 SOURCE Pristionchus pacificus
 ORGANISM Pristionchus pacificus
 Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;
 Neodiplogasteridae; Pristionchus.
 1 (bases 1 to 510)

REFERENCE

AUTHORS Srinivasan, J., Sinz, W., Lanz, C., Brand, A., Nandakumar, R.,
 Raddatz, G., Witte, H., Keller, H., Kipping, I., Pires da Silva, A.,
 Jesse, T., Millare, J., de Both, M., Schuster, S.C. and Sommer, R.J.
 A BAC-based genetic linkage map of the nematode Pristionchus
 pacificus
 Unpublished (2002)
 Contact: Sommer RJ
 Evolutionary Biology
 Max-Planck-Institute for Developmental Biology
 Spemannstr. 37-39, Tuebingen D-72076, Germany
 Tel: 00497071601371
 Fax: 00497071601498
 Email: ralf.sommer@uebingen.mpg.de
 Class: BAC ends.

FEATURES

source
 1..510
 Location/Qualifiers
 /organism="Pristionchus pacificus"
 /mol_type="genomic DNA"
 /strain="var. California"
 /db_xref="taxon:54126"
 /clone_lib="Pristionchus pacificus BAC ends"

ORIGIN

Query Match 48.1%; Score 26; DB 9; Length 510;
 Best Local Similarity 70.0%; Pred. No. 75;
 Matches 35; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 1 GATCTGATTGATTGAACGAAAAACGGAAAGCGTTTCAAGATAAATGCCA 50
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 253 GATCATATTGATTAAATGAAAATCGGGGAACGGGATATTTCGAAATGAGA 302

Search completed: January 18, 2006, 11:40:45
Job time : 232.555 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2006 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 17, 2006, 18:06:53 ; Search time 6461.19 Seconds
(without alignments)
11032.288 Million cell updates/sec

Title: US-10-511-327-2
Perfect score: 1254
Sequence: 1 gatctgattgattgaacgaa.....ttcgttcaatcaatcagatc 1254

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl.*

- 1: gb_ba.*
- 2: gb_in.*
- 3: gb_env.*
- 4: gb_on.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pr.*
- 9: gb_ro.*
- 10: gb_sts.*
- 11: gb_sy.*
- 12: gb_un.*
- 13: gb_vi.*
- 14: gb_htg.*
- 15: gb_pl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|--------|-------------|--------|----|-------------------------------|
| C 1 | 1147.2 | 91.5 | 3400 | 6 | AR477289 Sequence |
| C 2 | 1142.8 | 91.1 | 6122 | 6 | AX554413 Sequence |
| C 3 | 1142.8 | 91.1 | 6122 | 6 | AX554422 Sequence |
| C 4 | 1142.8 | 91.1 | 6122 | 6 | AX642149 Sequence |
| C 5 | 1142.8 | 91.1 | 6125 | 6 | AX554420 Sequence |
| C 6 | 1135.2 | 90.5 | 3399 | 11 | CVU46018 U46018 Cloning vec |
| C 7 | 1135.2 | 90.5 | 4100 | 6 | CS119871 Sequence |
| C 8 | 1131.8 | 90.3 | 1808 | 11 | AY781408 Synthetic |
| C 9 | 1130.8 | 90.2 | 3417 | 11 | AF153422 Cloning v |
| C 10 | 1129.6 | 90.1 | 1851 | 1 | ACCTN2670L M37690 Acinetobact |
| C 11 | 1129.6 | 90.1 | 1874 | 1 | ACCACEAA M62822 A.baumannii |
| C 12 | 1129.6 | 90.1 | 6464 | 1 | AY608912 Escherich |
| C 13 | 1129.6 | 90.1 | 11925 | 11 | AY303237 Shuttle v |
| C 14 | 1129.6 | 90.1 | 13515 | 11 | AY303236 Shuttle v |
| C 15 | 1129.6 | 90.1 | 45325 | 1 | AY123253 Klebsiell |
| C 16 | 1129.6 | 90.1 | 94281 | 1 | AP000342 Plasmid R |
| C 17 | 1129.6 | 90.1 | 120592 | 3 | AJ851089 Unculture |
| C 18 | 1129.6 | 90.1 | 218160 | 1 | AL513383 Salmonell |

| | | | | | |
|------|--------|------|-------|----|------------------------------|
| C 19 | 1127.4 | 89.9 | 5593 | 11 | AF519766 Cloning v |
| C 20 | 1127.2 | 89.9 | 7830 | 11 | AY428809 Cloning v |
| C 21 | 1125.2 | 89.7 | 2933 | 11 | AY781400 Synthetic |
| C 22 | 1122.6 | 89.5 | 4376 | 1 | U81140 Shigella fl |
| C 23 | 1122.6 | 89.5 | 6715 | 1 | AF326777 Shigella |
| C 24 | 1121.8 | 89.5 | 1319 | 11 | AY781404 Synthetic |
| C 25 | 1118.8 | 89.2 | 8530 | 11 | AY289598 Cloning v |
| C 26 | 1118.6 | 89.2 | 2701 | 11 | AF060240 Mobile an |
| C 27 | 1116.8 | 89.1 | 2839 | 11 | AY22812 Cloning v |
| C 28 | 1116.8 | 89.1 | 4152 | 11 | AY219693 Shuttle v |
| C 29 | 1116.8 | 89.1 | 5498 | 11 | AY222821 Shuttle v |
| C 30 | 1116 | 89.0 | 3930 | 11 | AB001326 Cloning v |
| C 31 | 1116 | 89.0 | 5240 | 11 | AB052891 Cloning v |
| C 32 | 1116 | 89.0 | 5273 | 11 | EVU51556 Expression |
| C 33 | 1116 | 89.0 | 6241 | 11 | AF092036 Shuttle v |
| C 34 | 1116 | 89.0 | 6758 | 11 | AF121784 Expressio |
| C 35 | 1116 | 89.0 | 8774 | 11 | IVU59267 Integratio |
| C 36 | 1116 | 89.0 | 9065 | 11 | AF405698 Reporter |
| C 37 | 1116 | 89.0 | 9071 | 11 | AB057644 Shuttle v |
| C 38 | 1116 | 89.0 | 12168 | 11 | AF405696 Reporter |
| C 39 | 1116 | 89.0 | 12448 | 11 | AF405697 Reporter |
| C 40 | 1116 | 89.0 | 14928 | 11 | AY230218 Expressio |
| C 41 | 1115.4 | 88.9 | 4907 | 11 | SYNBR328V L08858 pBR328 clon |
| C 42 | 1115.4 | 88.9 | 5996 | 11 | SYNBR325V L08855 pBR325 clon |
| C 43 | 1115 | 88.9 | 3475 | 11 | D88215 Cloning vec |
| C 44 | 1114.4 | 88.9 | 4589 | 11 | SYN261CAT M29363 Plasmid pFL |
| C 45 | 1114.4 | 88.9 | 5838 | 6 | AR072535 Sequence |

ALIGNMENTS

| | | | | | | | |
|------------|-------------------------------------|------------------------------------|------------------------------------|---------|-----|--------|-----------------|
| RESULT 1 | AR477289/c | AR477289 | Sequence 1 from patent US 6696278. | 3400 bp | DNA | linear | PAT 14-MAY-2004 |
| LOCUS | AR477289 | Sequence 1 from patent US 6696278. | | | | | |
| DEFINITION | AR477289 | | | | | | |
| ACCESSION | AR477289.1 | GI:47234624 | | | | | |
| VERSION | | | | | | | |
| KEYWORDS | | | | | | | |
| SOURCE | Unknown. | | | | | | |
| ORGANISM | Unclassified. | | | | | | |
| REFERENCE | 1 (bases 1 to 3400) | | | | | | |
| AUTHORS | Carstens, C.-P. | | | | | | |
| TITLE | Method for transfer of DNA segments | | | | | | |
| JOURNAL | Patent: US 6696278-A 1 24-FEB-2004; | | | | | | |
| FEATURES | Stratagene; La Jolla, CA | | | | | | |
| source | Location/Qualifiers | | | | | | |
| | 1..3400 | | | | | | |
| | /organism="unknown" | | | | | | |
| | /mol_type="genomic DNA" | | | | | | |

| | | | | | |
|--------|-----------------------|---|---------------------|----------------|--------------|
| ORIGIN | Query Match | 91.5% | Score 1147.2; | DB 6; | Length 3400; |
| | Best Local Similarity | 98.5% | Pred. No. 6.6e-309; | | |
| | Matches 1158; | Conservative | 0; | Mismatches 18; | Indels 0; |
| Qy | 50 | AAAACGGATCCTATCGTCAATTATTACCTCACGGGAGAGCCCTGAGCAAACTGGCCTCA | 109 | | |
| Db | 3053 | AAAAGGATCATATCGTCAATTATTACCTCACGGGAGAGCCCTGAGCAAACTGGCCTCA | 2994 | | |
| Qy | 110 | GGCATTTTGGAGACACAGGTCACACTGCTTCCTCGGTAGTCAATAAACCGGTAAACACGCA | 169 | | |
| Db | 2993 | GGCATTTTGGAGACACAGGTCACACTGCTTCCTCGGTAGTCAATAAACCGGTAAACACGCA | 2934 | | |
| Qy | 170 | ATAGACATAAGCGGCTATTAAACGACCGCTGCTTGAACCGACGACCGGGTTCGATTTGCT | 229 | | |
| Db | 2933 | ATAGACATAAGCGGCTATTAAACGACCGCTGCTTGAACCGACGACCGGGTTCGATTTGCT | 2874 | | |
| Qy | 230 | TTGGAATTTTGGCAATTCATCCGCTTATTATCACTTATTCAGGCGTAGCAACAGCGGTTT | 289 | | |
| Db | 2873 | TTGGAATTTTGGCAATTCATCCGCTTATTATCACTTATTCAGGCGTAGCAACAGCGGTTT | 2814 | | |


```
Db 1161 TTCTTACGGTCTTTAAAGGCGGTATATATCCAGCTGAAACGGTCTGTTATAGTACAT 1102
Qy 890 TGAGCAACTGACTGAAATGCCTCAAAATGTTCTTTACGATGCCATGGGATATATCAACG 949
Db 1101 TGAGCAACTGACTGAAATGCCTCAAAATGTTCTTTACGATGCCATGGGATATATCAACG 1042
Qy 950 GTGGTATATCCAGTGAATTTTTTCTCCAAATTTAGCTTCCCTAGCTCCGTAATAATCTCGAC 1009
Db 1041 GTGGTATATCCAGTGAATTTTTTCTCCAAATTTAGCTTCCCTAGCTCCGTAATAATCTCGAC 982
Qy 1010 AACTCAAAAAATACGCCGGTAGTGATCTTATTTCAATATGTTGAAAGTTCGAAACCTCTT 1069
Db 981 AACTCAAAAAATACGCCGGTAGTGATCTTATTTCAATATGTTGAAAGTTCGAAACCTCTT 922
Qy 1070 ACGTGCCGATCAACGCTCTCATTTTTGCGCAAAAGTTGCGCCAGGGCTTCCCGGTATCAACA 1129
Db 921 ACGTGCCGATCAACGCTCTCATTTTTGCGCAAAAGTTGCGCCAGGGCTTCCCGGTATCAACA 862
Qy 1130 GGGACACAGGATTTATTTATTCGCGAAGTGATCTTCCGTCACAGGTATTTATTCGGTC 1189
Db 861 GGGACACAGGATTTATTTATTCGCGAAGTGATCTTCCGTCACAGGTATTTATTCGGTC 802
Qy 1190 GAAAAG 1195
Db 801 GAAAAG 796

RESULT 3
AX554422/c
LOCUS AX554422 6122 bp DNA linear PAT 27-NOV-2002
DEFINITION Sequence 10 from Patent WO0246436.
ACCESSION AX554422
VERSION AX554422.1 GI:25898199
KEYWORDS synthetic construct
SOURCE synthetic construct
ORGANISM other sequences; artificial sequences.
REFERENCE
1 Bowdich, K.S., Barbas-Frederickson, S., Wild, M. and McWhirter, J.
TITILE Novel plasmid vectors
JOURNAL Patent: WO 0246436-A 10 13-JUN-2002;
ALEXION PHARMACEUTICALS, INC. (US)
FEATURES
Location/Qualifiers
1. 6122
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="plasmid"
ORIGIN
Query Match 91.1%; Score 1142.8; DB 6; Length 6122;
Best Local Similarity 99.8%; Pred. No. 1.1e-307;
Matches 1144; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 50 AAAACGGATCTATCGTCAATTTATACCTCCACGGGAGAGCCTGAGCAAACTGCGCTCA 109
Db 1941 AAAAAGGATCATATCGTCAATTTATACCTCCACGGGAGAGCCTGAGCAAACTGCGCTCA 1882
Qy 110 GGCAATTTGAGAAGCACACGGTCACACTGCTTCCGGTAGTCAATAAACCAGTAAACACAGCA 169
Db 1881 GGCAATTTGAGAAGCACACGGTCACACTGCTTCCGGTAGTCAATAAACCAGTAAACACAGCA 1822
Qy 170 ATAGACATAAGGGCTATTTAACGACCTGCGCTGAAACGGACGACCGGTCGAATTTGCT 229
Db 1821 ATAGACATAAGGGCTATTTAACGACCTGCGCTGAAACGGACGACCGGTCGAATTTGCT 1762
Qy 230 TTGGAATTTCTGCCATTCATCGCTTATTTACCTTATTCAGGCGTAGCAACACGAGCGTT 289
Db 1761 TTGGAATTTCTGCCATTCATCGCTTATTTACCTTATTCAGGCGTAGCAACACGAGCGTT 1702
Qy 290 TAAGGGCACCAATAACCTGCTTAAAAAAATACGCGCCGCTGCCACTCATCGCAGTAC 349
Db 1701 TAAGGGCACCAATAACCTGCTTAAAAAAATACGCGCGGCTGCCACTCATCGCAGTAC 1642
```

```
Qy 350 TGTGTAAATTCATTTAAGCATTTCTCCGACATGGAAGCCATCACAACCGGCATGATGAACC 409
Db 1641 TGTGTAAATTCATTTAAGCATTTCTCCGACATGGAAGCCATCACAACCGGCATGATGAACC 1582
Qy 410 TGAATCCCGAGCGGCATCAGCACCTTGTGCGCTTGGTATATAATTTTGGCCCATGGTGAAA 469
Db 1581 TGAATCCCGAGCGGCATCAGCACCTTGTGCGCTTGGTATATAATTTTGGCCCATGGTGAAA 1522
Qy 470 ACGGGGCGAAGAGTTGTCTCATATTGGCCACGTTTAAATCAAACTCGTGAAACTCAACC 529
Db 1521 ACGGGGCGAAGAGTTGTCTCATATTGGCCACGTTTAAATCAAACTCGTGAAACTCAACC 1462
Qy 530 CAGGGATTTGGCTGAGACGAAAAACATATTTCTCAATAAACCCCTTTAGGGAATATAGCCAGG 589
Db 1461 CAGGGATTTGGCTGAGACGAAAAACATATTTCTCAATAAACCCCTTTAGGGAATATAGCCAGG 1402
Qy 590 TTTTTCACCGTAACACGCGCATCTTGGCAATATATGTTAGAACTCGCGGAATTCGTCG 649
Db 1401 TTTTTCACCGTAACACGCGCATCTTGGCAATATATGTTAGAACTCGCGGAATTCGTCG 1342
Qy 650 TGGTATTTCACTCCAGAGCGATGAAAAAGTTTTCAGTTTGTCTCATGAAAAACGGTGTAACA 709
Db 1341 TGGTATTTCACTCCAGAGCGATGAAAAAGTTTTCAGTTTGTCTCATGAAAAACGGTGTAACA 1282
Qy 710 GGGTGAACACTATATCCCATATCACCAGCTCACCGTCTTTCATTGTCATACGTAATTCGGA 769
Db 1281 GGGTGAACACTATATCCCATATCACCAGCTCACCGTCTTTCATTGTCATACGTAATTCGGA 1222
Qy 770 TGACATTTATCAGCGGGGCAAGATGTGAATAAAGCGCGGATAAAACTTGTGCTTATTT 829
Db 1221 TGACATTTATCAGCGGGGCAAGATGTGAATAAAGCGCGGATAAAACTTGTGCTTATTT 1162
Qy 830 TTCTTTACGGTCTTTTAAAAAGGCGGTAATATCCAGCTGAAACGGTCTGGTTATAGGTACAT 889
Db 1161 TTCTTTACGGTCTTTTAAAAAGGCGGTAATATCCAGCTGAAACGGTCTGGTTATAGGTACAT 1102
Qy 890 TGACCAACTGACTGAAATGCTCAAAATGTTCTTTACGATGCCATTTGGGATATATCAACG 949
Db 1101 TGACCAACTGACTGAAATGCTCAAAATGTTCTTTACGATGCCATTTGGGATATATCAACG 1042
Qy 950 GTGGTATATCCAGTGAATTTTTTCTCCATTTAGCTTCCCTAGCTCCGTAATAATCTCGAC 1009
Db 1041 GTGGTATATCCAGTGAATTTTTTCTCCATTTAGCTTCCCTAGCTCCGTAATAATCTCGAC 982
Qy 1010 AACTCAAAAAATACGCCGGTAGTGATCTTATTTCAATATGTTGTAAGTTTGGAACTCTT 1069
Db 981 AACTCAAAAAATACGCCGGTAGTGATCTTATTTCAATATGTTGTAAGTTTGGAACTCTT 922
Qy 1070 ACGTGCCGATCAACGCTCTCATTTTTGCGCAAAAGTTGCGCCAGGGCTTCCCGGTATCAACA 1129
Db 921 ACGTGCCGATCAACGCTCTCATTTTTGCGCAAAAGTTGCGCCAGGGCTTCCCGGTATCAACA 862
Qy 1130 GGGACACAGGATTTATTTATTCGCGAAGTGATCTTCCGTCACAGGTATTTATTCGGTC 1189
Db 861 GGGACACAGGATTTATTTATTCGCGAAGTGATCTTCCGTCACAGGTATTTATTCGGTC 802
Qy 1190 GAAAAG 1195
Db 801 GAAAAG 796

RESULT 4
AX642149/c
LOCUS AX642149 6122 bp DNA linear PAT 21-FEB-2003
DEFINITION Sequence 1 from Patent WO0246435.
ACCESSION AX642149
VERSION AX642149.1 GI:28474637
KEYWORDS synthetic construct
SOURCE synthetic construct
ORGANISM other sequences; artificial sequences.
REFERENCE 1
```


AUTHORS Bowdish,K.S., Barbas-Frederickson,S., Lin,Y.C., Renshaw,M., Wild,M.
and McWhirter,J.
TITLE Engineered plasmids and their use for in situ production of genes
JOURNAL Patent: WO 0246435-A 1 13-JUN-2002;
ALEXION PHARMACEUTICALS, INC. (US)
FEATURES Location/Qualifiers
source 1. .6122
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="vector"

ORIGIN

Query Match 91.1%; Score 1142.8; DB 6; Length 6122;
Best Local Similarity 99.8%; Pred. No. 1.1e-307;
Matches 1144; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 50 AAAACGGATCTATCGTCAATATTATACCTCCACGGGAGAGCCTGAGCAAACTGGCCTCA 109
Db 1941 AAAAAGGATCATATCGTCAATATTATACCTCCACGGGAGAGCCTGAGCAAACTGGCCTCA 1882

Qy 110 GGCAATTTGAGAAGACACAGGTTCACCTGCTTCGGTAGTCAATAAACCAGGTAACACGCA 169
Db 1881 GGCATTTTGAGAAGACACAGGTTCACCTGCTTCGGTAGTCAATAAACCAGGTAACACGCA 1822

Qy 170 ATAGACATAAGCGGCTATTAAACGACCTCGCTGAAACGACGACCGGTCGAATTTGCT 229
Db 1821 ATAGACATAAGCGGCTATTAAACGACCTCGCTGAAACGACGACCGGTCGAATTTGCT 1762

Qy 230 TTGCAATTTTCGCCATTTCATCGCTTATTACCTTATTCAGGGGTAGCAACACGAGGGTT 289
Db 1761 TTGCAATTTTCGCCATTTCATCGCTTATTACCTTATTCAGGGGTAGCAACACGAGGGTT 1702

Qy 290 TAAGGGACCAATTAACCTGCTTAAAAAATTAACGACCTCGCTGAAACGACGACCGGTCGAATTTGCT 349
Db 1701 TAAGGGACCAATTAACCTGCTTAAAAAATTAACGACCTCGCTGAAACGACGACCGGTCGAATTTGCT 1642

Qy 350 TGTGTAATTCATTAAGCATTCGCGACATGGAAGCCATCAAAACGGCATGATGAAC 409
Db 1641 TGTGTAATTCATTAAGCATTCGCGACATGGAAGCCATCAAAACGGCATGATGAAC 1582

Qy 410 TGAATCGCAGCGGCATCAGACCTTTGCGCTTTGCGTATATATTTGCCCATGTTGAAA 469
Db 1581 TGAATCGCAGCGGCATCAGACCTTTGCGCTTTGCGTATATATTTGCCCATGTTGAAA 1522

Qy 470 ACGGGGGGAAGAGTTGTTCATATGCGCACGTTTAAATCAAAACTGGTGAAACTCAC 529
Db 1521 ACGGGGGGAAGAGTTGTTCATATGCGCACGTTTAAATCAAAACTGGTGAAACTCAC 1462

Qy 530 CAGGATTTGGCTGAGACGAAAAACATATTCATTAACACCTTTAGGGAATAGCCAGG 589
Db 1461 CAGGATTTGGCTGAGACGAAAAACATATTCATTAACACCTTTAGGGAATAGCCAGG 1402

Qy 590 TTTTCACCGTAAACAGCCACATCTTTGCGAATATATGTAGAAACTGCGGAAATCGTCG 649
Db 1401 TTTTCACCGTAAACAGCCACATCTTTGCGAATATATGTAGAAACTGCGGAAATCGTCG 1342

Qy 650 TGGTATTCACTCCAGAGCGATGAAAAAGTTTCAGTTTGTCTCATGGAATAACCGGTGAACAA 709
Db 1341 TGGTATTCACTCCAGAGCGATGAAAAAGTTTCAGTTTGTCTCATGGAATAACCGGTGAACAA 1282

Qy 710 GGGTGACACTATCCCATATCAGAGCTCAGCTTTTCATTTGCGCATACGTAATTCGGA 769
Db 1281 GGGTGACACTATCCCATATCAGAGCTCAGCTTTTCATTTGCGCATACGTAATTCGGA 1222

Qy 770 TGAGCAATTCATCAGCGGGCAGAGATGTGAATAAGCGCGGATAAAACTTGCTGCTATTT 829
Db 1221 TGAGCAATTCATCAGCGGGCAGAGATGTGAATAAGCGCGGATAAAACTTGCTGCTATTT 1162

Qy 830 TTCTTTACGGTCTTTAAAAAGCGCGTAAATATCCAGCTGAAACGGTCTGGTTATAGGTAC 889
Db 1161 TTCTTTACGGTCTTTAAAAAGCGCGTAAATATCCAGCTGAAACGGTCTGGTTATAGGTAC 1102

Qy 890 TGACCAACTCAGTGAAATGCTCAAAATGTTCTTTAGCATGCCATTCGGATATATCAACG 949
Db 1101 TGACCAACTCAGTGAAATGCTCAAAATGTTCTTTAGCATGCCATTCGGATATATCAACG 1042

Qy 950 GTGGTATATCAGTGATTTTCTCCATTTAGTCTCTTGTAGTCTCTGAAAAATCTCGAC 1009
Db 1041 GTGGTATATCAGTGATTTTCTCCATTTAGTCTCTTGTAGTCTCTGAAAAATCTCGAC 982

Qy 1010 AACTCAAAAATACGCCCGGTAGTGATCTTATTTCAATTATGGTGAAGTTGGAACTCTTT 1069
Db 981 AACTCAAAAATACGCCCGGTAGTGATCTTATTTCAATTATGGTGAAGTTGGAACTCTTT 922

Qy 1070 ACGTGCCGATCAACGCTCTCATTTTCGCAAAAGTTGCCACGAGGCTTCCCGGTATCAACA 1129
Db 921 ACGTGCCGATCAACGCTCTCATTTTCGCAAAAGTTGCCACGAGGCTTCCCGGTATCAACA 862

Qy 1130 GGGACACACAGATTTATTTATTTCTGGAAGTGATCTTCCGTACAGGTATTTATTCGGTC 1189
Db 861 GGGACACACAGATTTATTTATTTCTGCAAGTGATCTTCCGTACAGGTATTTATTCGGTC 802

Qy 1190 GAAAAG 1195
Db 801 GAAAAG 796

RESULT 5
AX554420/c
LOCUS AX554420 6125 bp DNA linear PAT 27-NOV-2002
DEFINITION Sequence 8 from Patent WO0246436.
ACCESSION AX554420
VERSION AX554420.1 GI:25898198
KEYWORDS
SOURCE synthetic construct
ORGANISM synthetic construct
other sequences; artificial sequences.

REFERENCE 1
AUTHORS Bowdish,K.S., Barbas-Frederickson,S., Wild,M. and McWhirter,J.
TITLE Novel plasmid vectors
JOURNAL Patent: WO 0246436-A 8 13-JUN-2002;
ALEXION PHARMACEUTICALS, INC. (US)
FEATURES Location/Qualifiers
source 1. .6125
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="plasmid"

ORIGIN

Query Match 91.1%; Score 1142.8; DB 6; Length 6125;
Best Local Similarity 99.8%; Pred. No. 1.1e-307;
Matches 1144; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 50 AAAACGGATCTATCGTCAATATTATACCTCCACGGGAGAGCCTGAGCAAACTGGCCTCA 109
Db 1941 AAAAAGGATCATATCGTCAATATTATACCTCCACGGGAGAGCCTGAGCAAACTGGCCTCA 1882

Qy 110 GGCAATTTGAGAAGACACAGGTTCACCTGCTTCGGTAGTCAATAAACCAGGTAACACGCA 169
Db 1881 GGCATTTTGAGAAGACACAGGTTCACCTGCTTCGGTAGTCAATAAACCAGGTAACACGCA 1822

Qy 170 ATAGACATAAGCGGCTATTAAACGACCTCGCTGAAACGACGACCGGTCGAATTTGCT 229
Db 1821 ATAGACATAAGCGGCTATTAAACGACCTCGCTGAAACGACGACCGGTCGAATTTGCT 1762

Qy 230 TTGCAATTTTCGCCATTTCATCGCTTATTACCTTATTCAGGGGTAGCAACACGAGGGTT 289
Db 1761 TTGCAATTTTCGCCATTTCATCGCTTATTACCTTATTCAGGGGTAGCAACACGAGGGTT 1702

Qy 290 TAAGGGACCAATTAACCTGCTTAAAAAATTAACGACCTCGCTGAAACGACGACCGGTCGAATTTGCT 349
Db 1701 TAAGGGACCAATTAACCTGCTTAAAAAATTAACGACCTCGCTGAAACGACGACCGGTCGAATTTGCT 1642

Qy 410 TGAATCGCAGCGGCATCAGACCTTTGCGCTTTGCGTATATATTTGCCCATGTTGAAA 469
Db 1581 TGAATCGCAGCGGCATCAGACCTTTGCGCTTTGCGTATATATTTGCCCATGTTGAAA 1522

Qy 470 ACGGGGGGAAGAGTTGTTCATATGCGCACGTTTAAATCAAAACTGGTGAAACTCAC 529
Db 1521 ACGGGGGGAAGAGTTGTTCATATGCGCACGTTTAAATCAAAACTGGTGAAACTCAC 1462

Qy 530 CAGGATTTGGCTGAGACGAAAAACATATTCATTAACACCTTTAGGGAATAGCCAGG 589
Db 1461 CAGGATTTGGCTGAGACGAAAAACATATTCATTAACACCTTTAGGGAATAGCCAGG 1402

Qy 590 TTTTCACCGTAAACAGCCACATCTTTGCGAATATATGTAGAAACTGCGGAAATCGTCG 649
Db 1401 TTTTCACCGTAAACAGCCACATCTTTGCGAATATATGTAGAAACTGCGGAAATCGTCG 1342

Qy 650 TGGTATTCACTCCAGAGCGATGAAAAAGTTTCAGTTTGTCTCATGGAATAACCGGTGAACAA 709
Db 1341 TGGTATTCACTCCAGAGCGATGAAAAAGTTTCAGTTTGTCTCATGGAATAACCGGTGAACAA 1282

Qy 710 GGGTGACACTATCCCATATCAGAGCTCAGCTTTTCATTTGCGCATACGTAATTCGGA 769
Db 1281 GGGTGACACTATCCCATATCAGAGCTCAGCTTTTCATTTGCGCATACGTAATTCGGA 1222

Qy 770 TGAGCAATTCATCAGCGGGCAGAGATGTGAATAAGCGCGGATAAAACTTGCTGCTATTT 829
Db 1221 TGAGCAATTCATCAGCGGGCAGAGATGTGAATAAGCGCGGATAAAACTTGCTGCTATTT 1162

Qy 830 TTCTTTACGGTCTTTAAAAAGCGCGTAAATATCCAGCTGAAACGGTCTGGTTATAGGTAC 889
Db 1161 TTCTTTACGGTCTTTAAAAAGCGCGTAAATATCCAGCTGAAACGGTCTGGTTATAGGTAC 1102

Db 1641 TGTGTGTAATTCATTAGCAATCTGCGCATGGAAGCCATCAAAACGGCATGATGAACC 1582
Qy 410 TGAATGCGCAGCGGCATCAGCACTTGTGCGCTTGGTATATAATTTGCCCATGTGTGAAA 469
Db 1581 TGAATGCGCAGCGGCATCAGCACTTGTGCGCTTGGTATATAATTTGCCCATGTGTGAAA 1522
Qy 470 ACGGGGCGAAGAGTTGTCCATATTTGGCCAGCTTTAAATCAAACTGGTGAACCTCACC 529
Db 1521 ACGGGGCGAAGAGTTGTCCATATTTGGCCAGCTTTAAATCAAACTGGTGAACCTCACC 1462
Qy 530 CAGGATTTGGCTGAGACGAAACATATTTCTCAATAAACCTTTAGGGAATAGGCCAGG 589
Db 1461 CAGGATTTGGCTGAGACGAAACATATTTCTCAATAAACCTTTAGGGAATAGGCCAGG 1402
Qy 590 TTTTCAACCGTAACAGCCACATCTTTGCGAATATATGTAGAACTCGCGGAAATCGTCG 649
Db 1401 TTTTCAACCGTAACAGCCACATCTTTGCGAATATATGTAGAACTCGCGGAAATCGTCG 1342
Qy 650 TGGTATTCACTCAGAGCGATGAAAACGTTTCAGTTTGTCTCATGAAAACGGGTGAACAA 709
Db 1341 TGGTATTCACTCAGAGCGATGAAAACGTTTTCAGTTTGTCTCATGAAAACGGGTGAACAA 1282
Qy 710 GGGTGAACACATATCCATATACAGCTCACCGCTCTTTTCATTGCCATACGTAATTCGGA 769
Db 1281 GGGTGAACACATATCCATATACAGCTCACCGCTCTTTTCATTGCCATACGTAATTCGGA 1222
Qy 770 TGAGCATTTCACTCAGCGGCGAAGATGTGAATAAAGGCGGATATAAACTTTGCTGCTATT 829
Db 1221 TGAGCATTTCACTCAGCGGCGAAGATGTGAATAAAGGCGGATATAAACTTTGCTGCTATT 1162
Qy 830 TTTCTTACGGTCTTTAAAGGCGGATATAATCCAGCTGAACGGTCTGTTATAGGTACAT 889
Db 1161 TTTCTTACGGTCTTTAAAGGCGGATATAATCCAGCTGAACGGTCTGTTATAGGTACAT 1102
Qy 890 TGAGCAACTGACTGAATGCTCAAAATGTTCTTTACGATGCGCATTTGGATATATCAACG 949
Db 1101 TGAGCAACTGACTGAATGCTCAAAATGTTCTTTACGATGCGCATTTGGATATATCAACG 1042
Qy 950 GTGGTATATCCAGTGAATTTTCTCCATTTTTCAGTCTCTTTCAGTCTCTGAAATCTCGAC 1009
Db 1041 GTGGTATATCCAGTGAATTTTCTCCATTTTTCAGTCTCTTTCAGTCTCTGAAATCTCGAC 982
Qy 1010 AACTCAAAATAATACCGCGGTAGTATCTTAATTTCAATATGTTGAAAGTTGGAACCTCTT 1069
Db 981 AACTCAAAATAATACCGCGGTAGTATCTTAATTTCAATATGTTGAAAGTTGGAACCTCTT 922
Qy 1070 ACGTCCGATCAACGTCTCATTTTTCGCAAAAGTTGGCCAGGCTTCCCGGTATCAACA 1129
Db 921 ACGTCCGATCAACGTCTCATTTTTCGCAAAAGTTGGCCAGGCTTCCCGGTATCAACA 862
Qy 1130 GGGACACAGGATTTATTATTCTCGGAAGTGATCTTCGGTCACAGGTATTATTTCGGTTC 1189
Db 861 GGGACACAGGATTTATTATTCTCGGAAGTGATCTTCGGTCACAGGTATTATTTCGGTTC 802
Qy 1190 GAAAG 1195
Db 801 GAAAG 796

RESULT 6

CVU46018/c 3399 bp DNA linear SYN 27-OCT-1999
LOCUS
DEFINITION Cloning vector pCRSCRIPT Cam, complete sequence.
ACCESSION U46018
VERSION U46018.1 GI:1184322
KEYWORDS Cloning vector pCRSCRIPT Cam
SOURCE Cloning vector pCRSCRIPT Cam
ORGANISM other sequences; artificial sequences; vectors.
REFERENCE 1 (bases 1 to 3399)
AUTHORS Marsh,S.
TITLE Direct Submission

JOURNAL
FEATURES
source

Submitted (16-JAN-1996) Sam Marsh, Marketing Analysis, Stratagene,
11011 North Torrey Pines Road, La Jolla, CA 92037, USA
Location/Qualifiers
1. 3399
/organism="Cloning vector pCRSCRIPT Cam"
/mol_type="genomic DNA"
/db_xref="taxon:45534"

ORIGIN

Query Match 90.5%; Score 1135.2; DB 11; Length 3399;
Best Local Similarity 98.4%; Pred. No. 1.5e-305;
Matches 1157; Conservative 0; Mismatches 18; Indels 1; Gaps 1;
Qy 50 AAAACGGATCTATCTGCTCAATTTATCTCTCAACGGGAGAGCTTGAGCAAACTGCGCTCA 109
Db 3052 AAAAAGGATCATATCGTCAATTTATCTCTCAACGGGAGAGCTTGAGCAAACTGCGCTCA 2993
Qy 110 GGCATTTTGAAGACACAGGTCACACTGCTTCCGGTAGTCAATAAACCGGTAAACACGCA 169
Db 2992 GGCATTTTGAAGACACAGGTCACACTGCTTCCGGTAGTCAATAAACCGGTAAACACGCA 2933
Qy 170 ATAGACATAACGGGCTATTAAACGACCTGCGCTGAAACCGACGACCGGGTCAAAATTTGCT 229
Db 2932 ATAGACATAACGGGCTATTAAACGACCTGCGCTGAAACCGACGACCGGGTCAAAATTTGCT 2873
Qy 230 TTGCAATTTTGCATTCATCCGCTTATTCACCTTATTCAGGCGTAGCAACCGGCGTT 289
Db 2872 TTGCAATTTTGCATTCATCCGCTTATTCACCTTATTCAGGCGTAGCAACCGGCGTT 2813
Qy 290 TAAGGGACCAATAAATGCTTAAATAAATTTACGCCGCGCTGCACTCATCGCAGTAC 349
Db 2812 TAAGGGACCAATAAATGCTTAAATAAATTTACGCCGCGCTGCACTCATCGCAGTAC 2753
Qy 350 TGTGTAATTCATTAAAGCATTCGCGACATGGAAGCCATCAAAACGGCATGATGAACC 409
Db 2752 TGTGTAATTCATTAAAGCATTCGCGACATGGAAGCCATCAAAACGGCATGATGAACC 2693
Qy 410 TGAATCGCCAGCGCATCAGCACCTTGTGCGCTTTCGGTATATAATTTGCCCATGCTGAAA 469
Db 2692 TGAATCGCCAGCGCATCAGCACCTTGTGCGCTTTCGGTATATAATTTGCCCATGCTGAAA 2633
Qy 470 ACGGGGCGAAGAGTTGTTCATATTTGGCCAGCTTTAAATCAAACTGGTGAACCTCACC 529
Db 2632 AC-GGGGCGAAGAGTTGTTCATATTTGCCACGTTTAAATCAAACTGGTGAACCTCACC 2574
Qy 530 CAGGATTTGGCTGAGACGAAACATATTTCTCAATAAACCCTTTAGGGAATAGGCCAGG 589
Db 2573 CAGGATTTGGCTGAGACGAAACATATTTCTCAATAAACCCTTTAGGGAATAGGCCAGG 2514
Qy 590 TTTTCAACCGTAACAGCCACATCTTTGCGAATATATGTAGAACTGCGCGAAATCGTCG 649
Db 2513 TTTTCAACCGTAACAGCCACATCTTTGCGAATATATGTAGAACTGCGCGAAATCGTCG 2454
Qy 650 TGGTATTCACTCAGAGCGATGAAAACGTTTCAGTTTGTCTCATGAAAACGGGTGAACAA 709
Db 2453 TGGTATTCACTCAGAGCGATGAAAACGTTTTCAGTTTGTCTCATGAAAACGGGTGAACAA 2394
Qy 710 GGGTGAACACTATCCCATATCACCAGCTCACCGCTTTTCATTGCGCATACGTAATTCGGA 769
Db 2393 GGGTGAACACTATCCCATATCACCAGCTCACCGCTTTTCATTGCGCATACGTAATTCGGA 2334
Qy 770 TGAGCATTTCACTCAGCGGCGAAGATGTGAATAAAGCCGGATAAACTTGTGCTTATTT 829
Db 2333 TGAGCATTTCACTCAGCGGCGAAGATGTGAATAAAGCCGGATAAACTTGTGCTTATTT 2274
Qy 830 TTTCTTACGGTCTTTAAAGGCGGCTAATATTCAGCTGAACGGTCTGGTTATAGGTACAT 889
Db 2273 TTTCTTACGGTCTTTAAAGGCGGCTAATATTCAGCTGAACGGTCTGGTTATAGGTACAT 2214
Qy 890 TGAGCAACTGACTGAATGCTCAAAATGTTCTTTACGATGCCATTTGGATATATCAACG 949
Db 2213 TGAGCAACTGACTGAATGCTCAAAATGTTCTTTACGATGCCATTTGGATATATCAACG 2154

```

Qy 950 GTGGTATATCAGTGATTTTTTTTCTCCATTTTAGTCTTCTAGCTCCCTGAAAACTCGAC 1009
Db 2153 GTGGTATATCAGTGATTTTTTTTCTCCATTTTAGTCTTCTAGCTCCCTGAAAACTCGAC 2094
Qy 1010 AACTCAAAAAATACGCCGGTAGTGATCTTATTTTCATTATGTTGAAAGTTGGAACCTCTT 1069
Db 2093 AACTCAAAAAATACGCCGGTAGTGATCTTATTTTCATTATGTTGAAAGTTGGAACCTCTT 2034
Qy 1070 ACGTGCCGATCAACGCTCTCATTTTTCGCAAAAGTTGGCCAGGGCTTCCCGGTATCAACA 1129
Db 2033 ACGTGCCGATCAACGCTCTCATTTTTCGCAAAAGTTGGCCAGGGCTTCCCGGTATCAACA 1974
Qy 1130 GGGACACAGGATTTATTTATTTCTGCGAAGTGATCTTCGGTCACAGGATTTATTTTCGTC 1189
Db 1973 GGGACACAGGATTTATTTATTTCTGCGAAGTGATCTTCCGTCACAGGATTTATTTTCGTC 1914
Qy 1190 GAAAAGGATCGGTTTTTCGCAATTTATCGTGAAACGCT 1225
Db 1913 GAAAAGGATCTAGGTGAAGATCCITTTTGATAATCT 1878

RESULT 7
CS119871/c CS119871 4100 bp DNA linear PAT 08-JUL-2005
LOCUS Sequence 133 from Patent WO2005056782.
ACCESSION CS119871
VERSION CS119871.1 GI:70667769
KEYWORDS
SOURCE
ORGANISM
synthetic construct
other sequences; artificial sequences.
REFERENCE
AUTHORS Amin,N.S., Boston,M.G., Bott,R.R., Cervin,M.A., Concar,E.M.,
Guatwiller,M.E., Jones,B.E., Liebeton,K., Miracle,G.S., Oh,H.,
Poulose,A.J., Raner,S.W., Scheibel,J.J., Weyler,W. and Whited,G.M.
Pethydrolase
TITLE Patent: WO 2005056782-A 133 23-JUN-2005;
JOURNAL GENENCOR INTERNATIONAL, INC. (US); THE PROCTER & GAMBLE COMPANY
(US)
FEATURES
source
Location/Qualifiers
1..4100
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="pAH502R plasmid"
ORIGIN
Query Match 90.5%; Score 1135.2; DB 6; Length 4100;
Best Local Similarity 98.4%; Pred. No. 1.5e-305;
Matches 1157; Conservative 0; Mismatches 18; Indels 1; Gaps 1;

Qy 50 AAAACGGATCTTCATCGTCAATTTATTTACCTCCACGGGAGAGCCTGAGCAAACTGGCCTCA 109
Db 3753 AAAAAGGATCATATCGTCAATTTATTTACCTCCACGGGAGAGCCTGAGCAAACTGGCCTCA 3694
Qy 110 GGCATTTGAGAAGCACAGGTCACACTGCTTCCGGTAGTCAATAAACCGGTAAACCCAGCA 169
Db 3693 GGCATTTGAGAAGCACACGGTCACACTGCTTCCGGTAGTCAATAAACCGGTAAACCCAGCA 3634
Qy 170 ATAGACATAAGCGGCTATTTAACGACCTCGCCCTGAACGACGACCGGGTCAAAATTTGCT 229
Db 3633 ATAGACATAAGCGGCTATTTAACGACCTCGCCCTGAACGACGACCGGGTCAAAATTTGCT 3574
Qy 230 TTCGAATTTCTGCCATTCATCGCTTATATCACTATTATTCAGGCGTAGCAACACGAGCGTT 289
Db 3573 TTCGAATTTCTGCCATTCATCGCTTATATCACTATTATTCAGGCGTAGCAACACGAGCGTT 3514
Qy 290 TAAGGGACCAATAACTGCGCTTAAAAAAATTAACGCCCGCCCTGCCACTCATCGAGTAC 349
Db 3513 TAAGGGACCAATAACTGCGCTTAAAAAAATTAACGCCCGCCCTGCCACTCATCGAGTAC 3454
Qy 350 TGTTCATTTCAATTAAGCAATTCGCGACATGGAAGCCATCACAACCGCATGATGAACC 409

```

```

Db 3453 TGTGTAAATTCATTAAGCATTTCTGCCGACATGGAAGCCATCACAACGCGATGATGAACC 3394
Qy 410 TGAATCCGACGCGCATCAGCACCTTGTGCGCTTGCCTTAATAATATTTGCCCATGGTGAAA 469
Db 3393 TGAATCCGACGCGCATCAGCACCTTGTGCGCTTGCCTTAATAATATTTGCCCATGGTGAAA 3334
Qy 470 ACGGGGCGAAGAGTTGTTCCATATGCGCCACGCTTTAAATCAAACTGGTGAAACTCAACC 529
Db 3333 AC-GGGGCGAAGAGTTGTTCCATATGCGCCACGCTTTAAATCAAACTGGTGAAACTCAACC 3275
Qy 530 CAGGGATTTGGCTGAGACGAAAAACATATTTCAATAAAACCTTTAGGAAATATGCCCAGG 589
Db 3274 CAGGGATTTGGCTGAGACGAAAAACATATTTCAATAAAACCTTTAGGAAATATGCCCAGG 3215
Qy 590 TTTTCACCGTAAACAGCGACATCTTTGGCAATATATGTTAGAAACTGCGGAAATTCGTG 649
Db 3214 TTTTCACCGTAAACAGCGACATCTTTGGCAATATATGTTAGAAACTGCGGAAATTCGTG 3155
Qy 650 TGGTATTTCACTCCAGAGCGATGAAAACGTTTTCAGTTTGCCTCATGAAAAACGTTGTAACA 709
Db 3154 TGGTATTTCACTCCAGAGCGATGAAAACGTTTTCAGTTTGCCTCATGAAAAACGTTGTAACA 3095
Qy 710 GGGTGAACACTATATCCCATATCACCAGCTCACCGCTTTTCATTTGCCATACGTAATTCGGGA 769
Db 3094 GGGTGAACACTATATCCCATATCACCAGCTCACCGCTTTTCATTTGCCATACGTAATTCGGGA 3035
Qy 770 TGACATTTCACTAGCGGGGCAAGATGTAATAAGGCCGATAAAACTTGTGCTTATTT 829
Db 3034 TGACATTTCACTAGCGGGGCAAGATGTAATAAGGCCGATAAAACTTGTGCTTATTT 2975
Qy 830 TTCTTTACGGTCTTTTAAAAAGGCGGTAAATATCCAGCTGAAACGGTCTGGTTATAGGTACAT 889
Db 2974 TTCTTTACGGTCTTTTAAAAAGGCGGTAAATATCCAGCTGAAACGGTCTGGTTATAGGTACAT 2915
Qy 890 TGACAAACTGACTGAAATGCTTCAAAATGTTCTTTACGATGCCATGGGATATATCAACG 949
Db 2914 TGACAAACTGACTGAAATGCTTCAAAATGTTCTTTACGATGCCATGGGATATATCAACG 2855
Qy 950 GTGTATATCAGTGATTTTTTTTCTCCATTTAGCTTCTTAGCTCTCTGAAAAATCTCGAC 1009
Db 2854 GTGTATATCAGTGATTTTTTTTCTCCATTTAGCTTCTTAGCTCTCTGAAAAATCTCGAC 2795
Qy 1010 AACTCAAAAAATACGCCGGTAGTGATCTTATTTTCATTATGTTGTAAGTTGGAACCTCTT 1069
Db 2794 AACTCAAAAAATACGCCGGTAGTGATCTTATTTTCATTATGTTGTAAGTTGGAACCTCTT 2735
Qy 1070 ACGTGCCGATCAACGCTCTCATTTTTCGCCAAAAGTTGGCCAGGGCTTCCCGGTATCAACA 1129
Db 2734 ACGTGCCGATCAACGCTCTCATTTTTCGCCAAAAGTTGGCCAGGGCTTCCCGGTATCAACA 2675
Qy 1130 GGGACACAGGATTTATTTCTGCGAAGTGATCTTCCGTCACAGGATTTATTTTCGGTC 1189
Db 2674 GGGACACAGGATTTATTTCTGCGAAGTGATCTTCCGTCACAGGATTTATTTTCGGTC 2615
Qy 1190 GAAAAGGATCGGTTTTTCGCAATTTATCGTGAAACGCT 1225
Db 2614 GAAAAGGATCTAGGTGAAGATCCITTTTGATAATCT 2579

RESULT 8
AV781408/c AV781408 1808 bp DNA linear SYN 28-FEB-2005
LOCUS Synthetic construct transposon mini-Mu transposon TncR1, complete
DEFINITION
sequence.
ACCESSION AV781408
VERSION AV781408.1 GI:60172112
KEYWORDS
SOURCE
ORGANISM
synthetic construct
synthetic construct
other sequences; artificial sequences.
REFERENCE
1 (bases 1 to 1808)
AUTHORS Zhang,C., Kitsberg,D., Chy,H., Zhou,Q. and Morrison,J.R.
TITLE Transposon-mediated generation of targeting vectors for the

```

production of gene knockouts
(er) Nucleic Acids Res. 33 (3), E24 (2005) In press
15699181
REFERENCE 2 (bases 1 to 1808)
AUTHORS Zhang, C.
TITLE Direct Submission
JOURNAL Submitted (14-OCT-2004) CopyRat Pty Ltd, 27-31 Wright Street,
Clayton, Victoria 3168, Australia
LOCATION/Qualifiers
1. .1808
/organism="synthetic construct"
/mol_type="other DNA"
/db_xref="taxon:32630"
repeat_region 1. .1808
/transposon="mini-Mu transposon TnCR1"
825. .1484
/codon_start=1
/transl_table=11
/product="chloramphenicol acetyltransferase"
/protein_id="AAK14475.1"
/db_xref="GI:60172113"
/translation="MEKKITGVTTVDISQMRKEHFEAFQSVACQNTQVQDITAF
LKTVMKHKFPAPFIHLRLMNAHPEFRMAKDGELVWDSVHPCVTVFHEOTET
SSLSEVHDDPROFLHIYSDVACGENLAVFKGFIENMFVSNPWSVTSFDLNV
ANMDFPAPVTMGKYYIQGDKVLMPLAIQVHVAVCDGFHVGRMLNELQYQDEWQGG
A"
ORIGIN
Query Match 90.3%; Score 1131.8; DB 11; Length 1808;
Best Local Similarity 98.0%; Pred. No. 1.3e-304;
Matches 1169; Conservative 0; Mismatches 17; Indels 7; Gaps 2;
QY 1 GATCTGATTGATGAACGAAACGCGAAGGTTTACGATATAATGCGAAACGGATCC 60
DB 1807 GATCTGAGCGCGCACGAAACGCGAAGGTTTACGATATAATGCGAAACGGCGG 1748
QY 61 TAT-----CGTCAATATTACTCCAGCGGAGAGCTGAGCAAACTGGCGTCAAGCAT 114
DB 1747 CCGCGCGCGCGTCAATATTACTCCAGCGGAGAGCTGAGCAAACTGGCGTCAAGCAT 1688
QY 115 TTGAGACGACACGCTGACACGCTTCGCGTAGTCAATAAACCGGTAAACGACCAATAGA 174
DB 1687 TTGAGACGACACGCTGACACGCTTCGCGTAGTCAATAAACCGGTAAACGACCAATAGA 1628
QY 175 CATAAGCGCTATTTAAGACCGCTCCCTGACCGACGACCGGTCGAAATTTGCTTTCGA 234
DB 1627 CATAAGCGCTATTTAAGACCGCTCCCTGACCGACGACCGGTCGAAATTTGCTTTCGA 1568
QY 235 ATTTCTGCCATTATCCGCTTATATCTATTATTCAGCGGTAGCAACGAGCGTTTAAGG 294
DB 1567 ATTTCTGCCATTATCCGCTTATATCTATTATTCAGCGGTAGC-ACCAGCGGTTTAAGG 1509
QY 295 GCACCAATACTGCTTAAAAAATATGCGCGCGCTGCGCACTCATCGCAGTACTGTG 354
DB 1508 GCACCAATACTGCTTAAAAAATATGCGCGCGCTGCGCACTCATCGCAGTACTGTG 1449
QY 355 TAAATTTAAGCATTTCTGCGACATGGAAGCATCACAAACGCGCATGATGAACCTGAAT 414
DB 1448 TAAATTTAAGCATTTCTGCGACATGGAAGCATCACAAACGCGCATGATGAACCTGAAT 1389
QY 415 CGCCAGCGCATCAGCACCTGTGCGCTTGCCTATATATTTTGGCCATGTTGAAACCGG 474
DB 1388 CGCCAGCGCATCAGCACCTGTGCGCTTGCCTATATATTTTGGCCATGTTGAAACCGG 1329
QY 475 GCGGAAGAGTTGTCCATATTGGCCACGTTTAAATCAAAACCTGTTGAACCTACCCAGG 534
DB 1328 GCGGAAGAGTTGTCCATATTGGCCACGTTTAAATCAAAACCTGTTGAACCTACCCAGG 1269
QY 535 ATTTGCTGAGACGAAACATATTCTCAATAAACCTTTAGGGAATAGGCCAGGTTTC 594
DB 1268 ATTTGCTGAGACGAAACATATTCTCAATAAACCTTTAGGGAATAGGCCAGGTTTC 1209
QY 595 ACCGTAACACGCGCATCTTGGCAATATATGTGTAGAAACTGCCGGAATCGTCGTGTA 654

Db 1208 ACCGTAACACGCGCATCTTCCGAATATATGTGTAGAAACTGCCGGAATCGTCGTGTA 1149
QY 655 TTCACTCCAGAGCGATGAAAAAGCTTTTCAGTTTGTCTCATGGAAGGCGTGTAAACAGGCGT 714
Db 1148 TTCACTCCAGAGCGATGAAAAAGCTTTTCAGTTTGTCTCATGGAAGGCGTGTAAACAGGCGT 1089
QY 715 AACACTATCCCATATACACAGCTCACCGTCTTTTCATTTGCCATAGCATTAATTCGGATGAGC 774
Db 1088 AACACTATCCCATATACACAGCTCACCGTCTTTTCATTTGCCATAGCGAATTCGGATGAGC 1029
QY 775 ATTCAATAGCGCGGCAAGATGTGAATAAAGCGCGGATGAAACCTTTGCTTATTTTCTT 834
Db 1028 ATTCAATAGCGCGGCAAGATGTGAATAAAGCGCGGATGAAACCTTTGCTTATTTTCTT 969
QY 835 TACGCTCTTTAAAAAGCGCGTAATATCCAGCTGAACCGTCTGTTTATAGGTACATTGAGC 894
Db 968 TACGCTCTTTAAAAAGCGCGTAATATCCAGCTGAACCGTCTGTTTATAGGTACATTGAGC 909
QY 895 AACTGACTGAATGCTCCAAAATGTTCTTTACGATGCCATTGGGATATATCAACGGTGT 954
Db 908 AACTGACTGAATGCTCCAAAATGTTCTTTACGATGCCATTGGGATATATCAACGGTGT 849
QY 955 ATATCCAGTGAATTTTCTCCATTTAGCTTCTTCTAGCTCCTGAAATCTCGCAACTC 1014
Db 848 ATATCCAGTGAATTTTCTCCATTTAGCTTCTTCTAGCTCCTGAAATCTCGCAACTC 789
QY 1015 AAAAAATACGCGCGTAGTGAATCTTATTTCAATATGTTGAAAGTTGGAACCTCTTACGTG 1074
Db 788 AAAAAATACGCGCGTAGTGAATCTTATTTCAATATGTTGAAAGTTGGAACCTCTTACGTG 729
QY 1075 CCGATCAACGCTCTCATTTTCGCCAAAAGTTGGCCAGGCTTCCCGGTATCAACAGGAC 1134
Db 728 CCGATCAACGCTCTCATTTTCGCCAAAAGTTGGCCAGGCTTCCCGGTATCAACAGGAC 669
QY 1135 ACCAGGATTTATTTATTTCTGCGAAGTGAATCTTCGTCACAGGATTTATTCGG 1187
Db 668 ACCAGGATTTATTTATTTCTGCGAAGTGAATCTTCGTCACAGGATTTATTCGG 616
RESULT 9
AF153422
LOCUS AF153422 3417 bp DNA circular SYN 02-APR-2000
DEFINITION Cloning vector pT8, complete sequence.
ACCESSION AF153422
VERSION AF153422.1 GI:7384995
KEYWORDS Cloning vector pT8
SOURCE Cloning vector pT8
ORGANISM Cloning vector pT8
REFERENCE 1 (bases 1 to 3417)
AUTHORS Yang, Y. and Spector, A.
TITLE Improved cloning vectors for transgene construction
JOURNAL Biotechniques 22 (6), 1032-1034 (1997)
PUBMED 9187746
REFERENCE 2 (bases 1 to 3417)
AUTHORS Yang, Y. and Spector, A.
TITLE Direct Submission
JOURNAL Submitted (21-MAY-1999) Institute of Molecular Biology, University
of Hong Kong, 8 Sassoon, Pokfulam, Hong Kong
FEATURES
source
1. .3417
/organism="Cloning vector pT8"
/mol_type="genomic DNA"
/db_xref="taxon:119062"
/note="derived from Strategene vector pBS KS(+) by
insertion of rare PstI and PacI restriction sites flanking
the multiple cloning site"
complement(241. .834)
/codon_start=1
/transl_table=11
/product="lacZ"
/protein_id="AAF61634.1"
CDS

```
/db_xref="GI:7384996"
/translation="MTMITPSAQAGQLTLTKGNSWVGPBPSRSTVSLISNSCSPG
DPLVLERPPHNSPSESYIILNARSIAVILORRWENPGVTQLNRLAHPFPAS
WRNSEARTDRPSQSLNSGWDAPCSGSAAGVVRVTSVATLASALAPAPFAF
PSFLATFAGFPRLNRLPLGFFRSALRHLDPKLD"
complement(2316..2867)
/codon_start=1
/transl_table=11
/product="chloramphenicol acetyltransferase"
/protein_id="AAF61635.1"
/db_xref="GI:7384997"
/translation="MEKKITGVTTVDISQHRKHFQSVQAOCTVQITVOLDITAF
LKTVKNNKHFVPAPIHILARLMAHPRLMAMKQDLVIWDSVHPCTVPHETOTF
SSLSEVHDDPROPLHIYSQDVACGENLAYFPKGFIEFMFVSANPNWVSFISDLNV
AMDNFFAPFSEWNIIRKATRC"
```

CDS

ORIGIN

```
Query Match      90.2%; Score 1130.8; DB 11; Length 3417;
Best Local Similarity 99.7%; Pred. No. 2.5e-304;
Matches 1143; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 50 AAAAGGATCTATCGTCAATTAATTAACCTCACGGGGAGAGCTGAGCAAACTGGCCCTCA 109
Db 1940 AAAAAGGATCATATCGTCAATTAATTAACCTCACGGGGAGAGCTGAGCAAACTGGCCCTCA 1999

QY 110 GGCATTTTCGAGAGCACACGGTCACACTGCTTCCGGTAGTCAATAAACCGGTAAACCCAGCA 169
Db 2000 GGCATTTTCGAGAGCACACGGTCACACTGCTTCCGGTAGTCAATAAACCGGTAAACCCAGCA 2059

QY 170 ATAGACATAAGCGGCTATTAAACGACCTCGCTCAACCGACGACCGGGTCAATTTGCT 229
Db 2060 ATAGACATAAGCGGCTATTAAACGACCTCGCTCAACCGACGACCGGGTCAATTTGCT 2119

QY 230 TTCGAATTTTCGCATTCATCGCTTATATCACTTTATTCAGGCGTAGCAACCGAGCGTT 289
Db 2120 TTCGAATTTTCGCATTCATCGCTTATATCACTTTATTCAGGCGTAGCAACCGAGCGTT 2179

QY 290 TAAGGGACCAATTAACCTGCTTAAAAAATTAACGCCCGCCCTGCCACTCATCGCAGTAC 349
Db 2180 TAAGGGACCAATTAACCTGCTTAAAAAATTAACGCCCGCCCTGCCACTCATCGCAGTAC 2239

QY 350 TGTTCGTAATTCATTAAAGCATTCGCGACATGGAAGCCATCAAAACGGCATGTAACCC 409
Db 2240 TGTTCGTAATTCATTAAAGCATTCGCGACATGGAAGCCATCAAAACGGCATGTAACCC 2299

QY 410 TGAATCGCAGGGCATCAGCACTTGTGCGCTTCGGTATATATATTTGCCCATGGTGAAG 469
Db 2300 TGAATCGCAGGGCATCAGCACTTGTGCGCTTCGGTATATATATTTGCCCATGGTGAAG 2359

QY 470 ACGGGGCGNAGAAGTTGTCATATTTGCCAGCTTTAAATCAAAACTGGTGAACCTCAC 529
Db 2360 AC-GGGGCGAAGAAGTTGTCATATTTGCCAGCTTTAAATCAAAACTGGTGAACCTCAC 2418

QY 530 CAGGATTTGGCTGAGACGAAAAACATATTTCTCAATAAACCCCTTTAGGAAATAGGCCAGG 589
Db 2419 CAGGATTTGGCTGAGACGAAAAACATATTTCTCAATAAACCCCTTTAGGAAATAGGCCAGG 2478

QY 590 TTTTCACCGTAACACGCCACATCTTTCCGATATATATGTGTAGAAACTCGCGAAATTCGTCG 649
Db 2479 TTTTCACCGTAACACGCCACATCTTTCCGATATATATGTGTAGAAACTCGCGAAATTCGTCG 2538

QY 650 TGGTATTCACCTCCAGAGCATGAAAACGGTTTCAGTTTCGTCTCATGAAAACCGTGTAAACA 709
Db 2539 TGGTATTCACCTCCAGAGCATGAAAACGGTTTCAGTTTCGTCTCATGAAAACCGTGTAAACA 2598

QY 710 GGGTGAACACTATCCATATACACAGCTCACCGCTCTTTTCATTGCGCATACGTAATTCGCGA 769
Db 2599 GGGTGAACACTATCCATATACACAGCTCACCGCTCTTTTCATTGCGCATACGTAATTCGCGA 2658

QY 770 TGAGCATTCATCAGCGGGCAAGAAATGTGAATAAAGCCGGGATAAAAACTTGTCCTTATTT 829
Db 2659 TGAGCATTCATCAGCGGGCAAGAAATGTGAATAAAGCCGGGATAAAAACTTGTCCTTATTT 2718
```

```
QY 830 TTCTTTACGGTCTTTTAAAAAGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACAT 889
Db 2719 TTCTTTACGGTCTTTTAAAAAGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACAT 2778

QY 890 TGAGCAACTGACTGAAATGCGCTCAAAATGTTCTTTTACGATGCCATTTGGGATATATCAACG 949
Db 2779 TGAGCAACTGACTGAAATGCGCTCAAAATGTTCTTTTACGATGCCATTTGGGATATATCAACG 2838

QY 950 GTGATATATCAGATGATTTTTTCTCCATTTTAGCTTCTTACGCTTCCGTAATAATCTCGAC 1009
Db 2839 GTGATATATCAGATGATTTTTTCTCCATTTTAGCTTCTTACGCTTCCGTAATAATCTCGAC 2898

QY 1010 AACTCAAAAAATACGCCGGTAGTGATCTTATTTTCAATTTATGTTGTAAGCTTGAACCTCTT 1069
Db 2899 AACTCAAAAAATACGCCGGTAGTGATCTTATTTTCAATTTATGTTGTAAGCTTGAACCTCTT 2958

QY 1070 ACGTGCCGATCAACGTCTCTCAATTTTTCGCAAAAGTTGCGCCAGGGCTTCCCGGTATCAACA 1129
Db 2959 ACGTGCCGATCAACGTCTCTCAATTTTTCGCAAAAGTTGCGCCAGGGCTTCCCGGTATCAACA 3018

QY 1130 GGGACACCAAGATTTATTTCTCGAAGTGATCTTCCGTCACAGGTATTTATTTTCGGTC 1189
Db 3019 GGGACACCAAGATTTATTTCTCGAAGTGATCTTCCGTCACAGGTATTTATTTTCGGTC 3078

QY 1190 GAAAAG 1195
Db 3079 GAAAAG 3084

RESULT 10
ACCTN2670L/c ACCTN2670L 1851 bp DNA linear BCT 01-MAR-1995
LOCUS Acinetobacter calcoaceticus Tn2670-like transposon encoding the CAT
DEFINITION structural gene, complete cds.
ACCESSION M37690
VERSION M37690.1 GI:141788
KEYWORDS CAT gene; Tn2670-like transposon.
SOURCE Acinetobacter calcoaceticus
ORGANISM Acinetobacter calcoaceticus
REFERENCE Moraxellaceae; Acinetobacter.
AUTHORS 1 (bases 1 to 1851)
Elisha, B.G. and Steyn, L.M.
TITLE Identification of an Acinetobacter baumannii gene region with
sequence and organizational similarity to Tn2670
JOURNAL Plasmid 25 (2), 96-104 (1991)
PUBMED 165008
COMMENT Original source text: Acinetobacter calcoaceticus anitratus DNA,
clone pGSH201.
Draft entry and computer-readable sequence for [Unpublished (1990)]
kindly submitted
by L.M.Steyn, 10-AUG-1990.
Univ of Cape Town
Dept Med Microbiol
Med Sch
Observatory
Cape Town 7925 South Africa.
Location/Qualifiers
1..1851
/organism="Acinetobacter calcoaceticus"
/mol_type="genomic DNA"
/isolate="SAK"
/sub_species="anitratus"
/db_xref="taxon:471"
/clone="pGSH201"
405..1064
/gene="CAT"
405..1064
/gene="CAT"
/codon_start=1
/transl_table=11
/protein_id="AAAG2571.1"
/db_xref="GI:141789"
FEATURES
source
1..1851
/organism="Acinetobacter calcoaceticus"
/mol_type="genomic DNA"
/isolate="SAK"
/sub_species="anitratus"
/db_xref="taxon:471"
/clone="pGSH201"
405..1064
/gene="CAT"
405..1064
/gene="CAT"
/codon_start=1
/transl_table=11
/protein_id="AAAG2571.1"
/db_xref="GI:141789"
gene
CDS
```

/translation="MEKKITGTTVDISQWHRKEHFEAFQSVACQTYNQTVDITAF
LKTVKNKHFPYPAFIHLARLNMNAHPEFRMAMKDELIVWDSVHPCYTVFHEQTET
SSLWSEYHDDFRQFLHIYSQDVACYGLENLAFPKGFLENMFVSNPWSVSTFDLNV
AMNDNFPAPVFTMGKYYTQGDVKVLMPLAIQVHHAVCDGFHVRMLNELQQYCDWEQGG
A"

ORIGIN

Query Match 90.1%; Score 1129.6; DB 1; Length 1851;
Best Local Similarity 98.8%; Pred. No. 5.4e-304;
Matches 1138; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

| | | | |
|----|------|--|------|
| QY | 50 | AAACCGGATCCTATCGTCAATTTATTTACCTCCACGGGGAGAGCCTCAGCAAACTGCGCTCA | 109 |
| DB | 1333 | ATAAATGATCATATCGTCAATTTATTTACCTCCACGGGGAGAGCCTCAGCAAACTGCGCTCA | 1274 |
| QY | 110 | GGCATTTTGAGAGACACAGGTCACTGCTTCCGGTGTAGTCAATAAACCAGGTAACACGCA | 169 |
| DB | 1273 | GGCATTTTGAGAGACACAGGTCACTGCTTCCGGTGTAGTCAATAAACCAGGTAACACGCA | 1214 |
| QY | 170 | ATAGACATAAGCGGTATTTAAGACCTCGCTGAACGACGACCGGGTGAATTTGCT | 229 |
| DB | 1213 | ATAGACATAAGCGGTATTTAAGACCTCGCTGAACGACGACCGGGTGAATTTGCT | 1154 |
| QY | 230 | TTGGAATTTGCGCAATTCATCGCTTATATCACTTATTCAGGCGTAGCAACAGGCGTT | 289 |
| DB | 1153 | TTGGAATTTGCGCAATTCATCGCTTATATCACTTATTCAGGCGTAGCAACAGGCGTT | 1094 |
| QY | 290 | TAAAGGACCAATAAATCGCTTAAATAAATAAGCCCGCTGCACTCATCGAGTAC | 349 |
| DB | 1093 | TAAAGGACCAATAAATCGCTTAAATAAATAAGCCCGCTGCACTCATCGAGTAC | 1034 |
| QY | 350 | TGTTGTAATTCATTAAGCATTCGCGACATGAAGCCATCAAAACGCGCATGTAAC | 409 |
| DB | 1033 | TGTTGTAATTCATTAAGCATTCGCGACATGAAGCCATCAAAACGCGCATGTAAC | 974 |
| QY | 410 | TGAATCGCAGCGGATCAGCACTTGTGCGCTTCGGTATATATTTGCCCATGTGAAA | 469 |
| DB | 973 | TGAATCGCAGCGGATCAGCACTTGTGCGCTTCGGTATATATTTGCCCATGTGAAA | 914 |
| QY | 470 | ACGGGGGCGAAGAGTTGTCCATATTTGCGCAGCTTTAAATCAAACTGTAACCTCA | 529 |
| DB | 913 | ACGGGGGCGAAGAGTTGTCCATATTTGCGCAGCTTTAAATCAAACTGTAACCTCA | 854 |
| QY | 530 | CAGGGAATTTGGCTGAGACGAAACATATTTCAATAAACCCTTTAGGGAATAGGCCA | 589 |
| DB | 853 | CAGGGAATTTGGCTGAGACGAAACATATTTCAATAAACCCTTTAGGGAATAGGCCA | 794 |
| QY | 590 | TTTTTACCCTTAACACGCCACATCTTTCGGAATATATGTGTAGAACTGCGGAAATCG | 649 |
| DB | 793 | TTTTTACCCTTAACACGCCACATCTTTCGGAATATATGTGTAGAACTGCGGAAATCG | 734 |
| QY | 650 | TGGTATTCACCTCAGACGAGTAAACCTTTTCAGTTTCTCATGTAACAAACCGTGTAA | 709 |
| DB | 733 | TGGTATTCACCTCAGACGAGTAAACCTTTTCAGTTTCTCATGTAACAAACCGTGTAA | 674 |
| QY | 710 | GGGTGAACACATATCCCATATCACAGCTCACCGTCTTTTCATTCGCATACGTAATTC | 769 |
| DB | 673 | GGGTGAACACATATCCCATATCACAGCTCACCGTCTTTTCATTCGCATACGTAATTC | 614 |
| QY | 770 | TGAGCATTCATCAGCGGGCAAGATGTGAATTAAGCGCGGATATAAACTTTGTCTTAT | 829 |
| DB | 613 | TGAGCATTCATCAGCGGGCAAGATGTGAATTAAGCGCGGATATAAACTTTGTCTTAT | 554 |
| QY | 830 | TTCTTTTACGGCTTTTAAAGCCGCTTAATTCACGCTGAACGGTCTGTTATAGTACAT | 889 |
| DB | 553 | TTCTTTTACGGCTTTTAAAGCCGCTTAATTCACGCTGAACGGTCTGTTATAGTACAT | 494 |
| QY | 890 | TGAGCAACTGACTGAAATGCCTCAAAATGTTCTTTTACGATGCAATTTGGGATATATCA | 949 |
| DB | 493 | TGAGCAACTGACTGAAATGCCTCAAAATGTTCTTTTACGATGCAATTTGGGATATATCA | 434 |
| QY | 950 | GTGGTATATCCAGTGAATTTTTTCTTCATTTTTCCTTTCCTTTCCTTTCCTTTCCT | 1009 |

| | | | |
|--|------|--|------|
| DB | 433 | GTGGTATATCCAGTGAATTTTTTCTCCATTTAGTCTCTTAGCTCCGAAATCTCGAT | 374 |
| QY | 1010 | AACTCAAAAAATAGCCCGGTAGTATCTTATTTCAATTTATGTTGAAAGTTTGAACCTCT | 1069 |
| DB | 373 | AACTCAAAAAATAGCCCGGTAGTATCTTATTTCAATTTATGTTGAAAGTTTGAACCTCT | 314 |
| QY | 1070 | ACGTGCCGATCAACGTCTCATTTTTCGCCAAAGTTGGCCAGGGTCTCCCGGTATCAACA | 1129 |
| DB | 313 | ACGTGCCGATCAACGTCTCATTTTTCGCCAAAGTTGGCCAGGGTCTCCCGGTATCAACA | 254 |
| QY | 1130 | GGGACACAGCATTTATTTATTTCTCGAAGTGATCTTCCGTACAGGTATTTATTCGGT | 1189 |
| DB | 253 | GGGACACAGCATTTATTTATTTCTCGAAGTGATCTTCCGTACAGGTATTTATTCGGG | 194 |
| QY | 1190 | GAAGAGGATCCG 1201 | |
| DB | 193 | CAAGTGGTGC 182 | |
| RESULT 11 | | | |
| ACCACEAA/c | | | |
| LOCUS | | | |
| DEFINITION A.baumannii chloramphenicol acetyltransferase (cat) gene, complete cds. | | | |
| ACCESSION M62822 | | | |
| VERSION M62822.1 GI:141744 | | | |
| KEYWORDS chloramphenicol acetyltransferase. | | | |
| SOURCE Acinetobacter baumannii | | | |
| ORGANISM Acinetobacter baumannii | | | |
| Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Moraxellaceae; Acinetobacter. | | | |
| REFERENCE 1 (bases 1 to 1874) | | | |
| AUTHORS Elisha,B.G. and Steyn,L.M. | | | |
| TITLE Identification of an Acinetobacter baumannii gene region with sequence and organizational similarity to Tn2670 | | | |
| JOURNAL Plasmid (1991) in press | | | |
| COMMENT Original source text: A.baumannii DNA. | | | |
| FEATURES | | | |
| source | | | |
| 1..1874 | | | |
| /organism="Acinetobacter baumannii" | | | |
| /mol_type="genomic DNA" | | | |
| /db_xref="taxon:470" | | | |
| gene | | | |
| 405..1064 | | | |
| /gene="CAT" | | | |
| 405..1064 | | | |
| /gene="CAT" | | | |
| /codon_start=1 | | | |
| /transl_table=1 | | | |
| /product="chloramphenicol acetyltransferase" | | | |
| /protein_id="AAA21891.1" | | | |
| /db_xref="GI:141745" | | | |
| /translation="MEKKITGTTVDISQWHRKEHFEAFQSVACQTYNQTVDITAF | | | |
| LKTVKNKHFPYPAFIHLARLNMNAHPEFRMAMKDELIVWDSVHPCYTVFHEQTET | | | |
| SSLWSEYHDDFRQFLHIYSQDVACYGLENLAFPKGFLENMFVSNPWSVSTFDLNV | | | |
| AMNDNFPAPVFTMGKYYTQGDVKVLMPLAIQVHHAVCDGFHVRMLNELQQYCDWEQGG | | | |
| A" | | | |
| ORIGIN | | | |
| Query Match 90.1%; Score 1129.6; DB 1; Length 1874; | | | |
| Best Local Similarity 98.8%; Pred. No. 5.4e-304; | | | |
| Matches 1138; Conservative 0; Mismatches 14; Indels 0; Gaps 0; | | | |
| QY | 50 | AAACCGGATCCTATCGTCAATTTATTTACCTCCACGGGGAGAGCCTGAGCAAACTGCGCTCA | 109 |
| DB | 1333 | ATAAATGATCATATCGTCAATTTATTTACCTCCACGGGGAGAGCCTGAGCAAACTGCGCTCA | 1274 |
| QY | 110 | GGCATTTTGAGAGACACAGGTCACTGCTTCCGGTGTAGTCAATAAACCAGGTAACACGCA | 169 |
| DB | 1273 | GGCATTTTGAGAGACACAGGTCACTGCTTCCGGTGTAGTCAATAAACCAGGTAACACGCA | 1214 |
| QY | 170 | ATAGACATAAGCGGTATTTAAGACCTCGCTGAACGACGACCGGGTGAATTTGCT | 229 |
| DB | 1213 | ATAGACATAAGCGGTATTTAAGACCTCGCTGAACGACGACCGGGTGAATTTGCT | 1154 |

| | | | |
|---|------|--|------|
| Qy | 230 | TTTGGAAATTTCTGCCATTCATCCGCTTATTATCACTTATTTCAGGCGTAGCAACACGAGCGTT | 289 |
| Db | 1153 | TTTGGAAATTTCTGCCATTCATCCGCTTATTATCACTTATTTCAGGCGTAGCAACACGAGCGTT | 1094 |
| Qy | 290 | TAAGGCGACCAATAACTGCTTTAAAAAATTAACGCCCGCCCTGCCACTCATCGCAGTAC | 349 |
| Db | 1093 | TAAGGCGACCAATAACTGCTTTAAAAAATTAACGCCCGCCCTGCCACTCATCGCAGTAC | 1034 |
| Qy | 350 | TGTTGTAAATTCATTAAGCAATTCGCGACATGGAAGCCATCACAAACGGCATGATGAACC | 409 |
| Db | 1033 | TGTTGTAAATTCATTAAGCAATTCGCGACATGGAAGCCATCACAAACGGCATGATGAACC | 974 |
| Qy | 410 | TGAATCGCCAGCGGCATCAGCACCTTGTGCGCTTGCCTGATATAATTTGCCCATGCTGAAA | 469 |
| Db | 973 | TGAATCGCCAGCGGCATCAGCACCTTGTGCGCTTGCCTGATATAATTTGCCCATGCTGAAA | 914 |
| Qy | 470 | ACGGGGCGAAGAGTGTGTCATATTGGCCACGTTTAAATCAAACTGGTGAACCTCACC | 529 |
| Db | 913 | ACGGGGCGAAGAGTGTGTCATATTGGCCACGTTTAAATCAAACTGGTGAACCTCACC | 854 |
| Qy | 530 | CAGGGAATGGCTGAGACGAAAAACATATCTCAATAAACCCCTTTAGGGAATAAGGCCAGG | 589 |
| Db | 853 | CAGGGAATGGCTGAGACGAAAAACATATCTCAATAAACCCCTTTAGGGAATAAGGCCAGG | 794 |
| Qy | 590 | TTTTTCAACCGTAAACAGCCACATCTTGCGAATATATGTGTAGAACTGCCGGAATCGTCG | 649 |
| Db | 793 | TTTTTCAACCGTAAACAGCCACATCTTGCGAATATATGTGTAGAACTGCCGGAATCGTCG | 734 |
| Qy | 650 | TGTTATTCACCTCAGAGCGGATGAAAAAGCTTTTCAGTTTGTCTATGGAAGCCGTGTAAACA | 709 |
| Db | 733 | TGTTATTCACCTCAGAGCGGATGAAAAAGCTTTTCAGTTTGTCTATGGAAGCCGTGTAAACA | 674 |
| Qy | 710 | GGGTGAACACTATCCCATATCACAGCTCACCGTCTTTTCATTTGCGCATACGTAATTCGGA | 769 |
| Db | 673 | GGGTGAACACTATCCCATATCACAGCTCACCGTCTTTTCATTTGCGCATACGTAATTCGGA | 614 |
| Qy | 770 | TGAGCATTCATCAGGCGGGCAAGATGTGAATAAGGCGCGATGAACTTGTGCTTATTT | 829 |
| Db | 613 | TGAGCATTCATCAGGCGGGCAAGATGTGAATAAGGCGCGATGAACTTGTGCTTATTT | 554 |
| Qy | 830 | TTCTTTACGCTCTTTAAAAAGCGCGTAATATCCAGCTGAACGCTCTGGTTATAGGTACAT | 889 |
| Db | 553 | TTCTTTACGCTCTTTAAAAAGCGCGTAATATCCAGCTGAACGCTCTGGTTATAGGTACAT | 494 |
| Qy | 890 | TGAGCAACTGACTGAAATGCCTCAAAATGTTCTTTACGATGCCATTTGGGATATATCAACG | 949 |
| Db | 493 | TGAGCAACTGACTGAAATGCCTCAAAATGTTCTTTACGATGCCATTTGGGATATATCAACG | 434 |
| Qy | 950 | GTGGTATATCCAGTGATTTTTTTTCTCCATTTAGCTTCCTTAGCTCCTGAAAACTCGAC | 1009 |
| Db | 433 | GTGGTATATCCAGTGATTTTTTTTCTCCATTTAGCTTCCTTAGCTCCTGAAAACTCGAT | 374 |
| Qy | 1010 | AACTCAAAAAATACGCCGCTAGTGATCTTTATTTTCATTTATGGTGAAGTTGGAACTCTTT | 1069 |
| Db | 373 | AACTCAAAAAATACGCCGCTAGTGATCTTTATTTTCATTTATGGTGAAGTTGGAACTCTTT | 314 |
| Qy | 1070 | ACGTGCCGATCAACGCTCTCATTTTCGCCAAAAGTTGGCCCGAGGCTTCCCGGTATCAACA | 1129 |
| Db | 313 | ACGTGCCGATCAACGCTCTCATTTTCGCCAAAAGTTGGCCCGAGGCTTCCCGGTATCAACA | 254 |
| Qy | 1130 | GGGACACCGAGTTTATTTATCTCGGAAGTGATCTTCGGTACAGGTATTTATTCGGTC | 1189 |
| Db | 253 | GGGACACCGAGTTTATTTATCTCGGAAGTGATCTTCGGTACAGGTATTTATTCGGTC | 194 |
| Qy | 1190 | GAAGAAGGATCCG | 1201 |
| Db | 193 | CAAAAGTGGCTCG | 182 |
| RESULT 12 | | | |
| AY608912 | | | |
| LOCUS | | | |
| AY608912 6464 bp DNA circular BCT 01-JUN-2004 | | | |

| | |
|--------------|---|
| DEFINITION | Escherichia coli plasmid pFL129, complete sequence. |
| ACCESSION | AY608912 |
| VERSION | AY608912.1 |
| KEYWORDS | GI:47717944 |
| SOURCE | Escherichia coli |
| ORGANISM | Escherichia coli |
| REFERENCE | Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia. |
| AUTHORS | 1 (bases 1 to 6464) |
| TITLE | Wild,J., Czyz,A., Rakowski,S.A. and Filutowicz,M. |
| JOURNAL | Mobilizable gamma ori plasmid pFL129 sequence |
| REFERENCE | 2 (bases 1 to 6464) |
| AUTHORS | Rakowski,S.A., Wild,J. and Filutowicz,M. |
| TITLE | Direct Submission |
| JOURNAL | Submitted (27-APR-2004) Bacteriology and Oncology, University of Wisconsin - Madison, 420 Henry Mall - Room 151, Madison, WI 53706, USA |
| FEATURES | Location/Qualifiers |
| source | 1..6464 |
| | /organism="Escherichia coli" |
| | /mol_type="genomic DNA" |
| | /db_xref="taxon:562" |
| | /plasmid="pFL129" |
| rep_origin | 1..383 |
| | /note="gamma origin of plasmid R6K" |
| enhancer | 1..106 |
| | /note="gamma ori enhancer" |
| misc_feature | 107..198 |
| | /note="A+T-rich region" |
| protein_bind | 199..352 |
| | /note="contains 22 bp direct repeat" |
| gene | 431..1348 |
| | /gene="pir" |
| CDS | 431..1348 |
| | /gene="pir" |
| | /note="replication initiator (Rep) for R6K" |
| | /codon_start=1 |
| | /transl_table=11 |
| | /product="pi" |
| | /protein_id="AAT37965.1" |
| | /db_xref="GI:47717946" |
| | /translation="MRLKVMVDNKKTKIRHNLNHTLAQLPLPAKRVVMYALAPID |
| | SKPLEGRVFKRAEDLAALAKITPSLAYROLKEGKLGASKISLRGDDILALAKE |
| | LNLPFTAKNSPEELDLNIIEMIAVNDSEGLSKFTIEPIYISLLGKKKFTQLL |
| | TASLRSSQYSSSLYLIRKHSNFKKNYFIISVDELKEELIATYTFDKDGNIEYKYP |
| | DFPIFKDLNKAIAEIKKTEISFVGFTVHEKEGRKISKLFKFEFVDEDEFSGDKDD |
| | EAPFMNLSEADAFLKVDETVPPKAG" |
| variation | 750 |
| | /gene="pir" |
| | /note="compared to plasmid pFL130; results in |
| | phenylalanine to serine substitution; predicted to produce |
| | a copy-up phenotype" |
| | /replace="C" |
| | /complement(2136..2795) |
| gene | /gene="cat" |
| | /complement(2136..2795) |
| CDS | /gene="cat" |
| | /note="chloramphenicol resistance marker" |
| | /codon_start=1 |
| | /transl_table=11 |
| | /product="chloramphenicol acetyltransferase" |
| | /protein_id="AAT37967.1" |
| | /db_xref="GI:47717948" |
| | /translation="MEKKITGYTTVDISQWHRKEHFEAFQSAOCTYNQTVOLDITAF |
| | LKTVKKNKHFYPAPFIHLARLNMNAHPEFRMAKDGELVWDSVHPCVTVPEQTET |
| | SSLMSEYHDDFRQFLHIYSQDVACYGENLAYFPKGFLENMFVSANPWVSTSDLVN |
| | ANMNFAPFTMGKTYTQDGKVLMLPLAIQVHHAVCDGFHVGRLNELQYCDSEWQGS |
| | A" |
| orit | 3692..3803 |
| | /note="from plasmid RK2 (RP4)" |
| gene | /complement(4180..5454) |

| Query Match | 90.1%; | Score 1129.6; | DB 1; | Length 6464; |
|-----------------------|-----------------|---|-----------|--------------|
| Best Local Similarity | 98.8%; | Pred. No. 5.6e-304; | | |
| Matches 1138; | Conservative 0; | Mismatches 14; | Indels 0; | Gaps 0; |
| Qy | 50 | AAACGGATCCTATCGTCAATATTACTCCACGGGGAGAGCCTGAGCAACCTGGGCTCA | 109 | |
| Db | 1867 | ATAAATGATCATATCGTCAATATTACTCCACGGGGAGAGCCTGAGCAACCTGGGCTCA | 1926 | |
| Qy | 110 | GGCATTGTGGAAGCACACGGGTCACTGCTTCGGGTAGTCAATAAACCAGGTAACACGCA | 169 | |
| Db | 1927 | GGCATTGTGGAAGCACACGGGTCACTGCTTCGGGTAGTCAATAAACCAGGTAACACGCA | 1986 | |
| Qy | 170 | ATAGACATAAGCGGCTATTTAAACGACCTCGCTGAAACCGACGACGGGTGCAATTTGGT | 229 | |
| Db | 1987 | ATAGACATAAGCGGCTATTTAAACGACCTCGCTGAAACCGACGACGGGTGCAATTTGGT | 2046 | |
| Qy | 230 | TTCCGAATTTCTGCCATTCATCCGCTTATTATCACTTTATTCAGCGGTAGCAACACGCGTT | 289 | |
| Db | 2047 | TTCCGAATTTCTGCCATTCATCCGCTTATTATCACTTTATTCAGCGGTAGCAACACGCGTT | 2106 | |
| Qy | 290 | TAAAGGCGACCAATAACTGCCCTTAAAAAATTAAGCCCGCCCTGCCACTCATCGCAGTAC | 349 | |
| Db | 2107 | TAAAGGCGACCAATAACTGCCCTTAAAAAATTAAGCCCGCCCTGCCACTCATCGCAGTAC | 2166 | |
| Qy | 350 | TGTTGTAAATTCATTAAGCATCTTGCCGACATGGAAGGCATCAAAACGGGATGATGAACC | 409 | |
| Db | 2167 | TGTTGTAAATTCATTAAGCATCTTGCCGACATGGAAGGCATCAAAACGGGATGATGAACC | 2226 | |
| Qy | 410 | TGAATCGCCAGCGGCATACGACCTTGTCGCTTGCGTATAAATTTGCCCATGGTGAAA | 469 | |
| Db | 2227 | TGAATCGCCAGCGGCATACGACCTTGTCGCTTGCGTATAAATTTGCCCATGGTGAAA | 2286 | |
| Qy | 470 | ACGGGGCGGAAGAGTTGTCCATATTTGGCCACGTTTAAATCAAAACTGGTGAAACTCACC | 529 | |
| Db | 2287 | ACGGGGCGGAAGAGTTGTCCATATTTGGCCACGTTTAAATCAAAACTGGTGAAACTCACC | 2346 | |
| Qy | 530 | CAGGGATTTGCTGAGACGAAAAACATATTTCTCAATAAACCTTTAGGGAAATAGGCCAGG | 599 | |
| Db | 2347 | CAGGGATTTGCTGAGACGAAAAACATATTTCTCAATAAACCTTTAGGGAAATAGGCCAGG | 2406 | |


```
Db 10948 TGGTATTCACTCCAGAGCGATGAACCGTTTCAGTTTCTCATGGAACCGGTGAACAA 11007
Qy 710 GGGTGAACATATCCCATATCACCAGCTCACCGTCTTTTCATTCGCATACGTAATTCGGA 769
Db 11008 GGGTGAACATATCCCATATCACCAGCTCACCGTCTTTTCATTCGCATACGTAATTCGGA 11067
Qy 770 TGAGCATTCATCAGCGGGCGGAGATGTAATAAGCGCGGATGAACCTTGCTTATT 829
Db 11068 TGAGCATTCATCAGCGGGCGGAGATGTAATAAGCGCGGATGAACCTTGCTTATT 11127
Qy 830 TTCTTTACGGTCTTTAAAAAGCGCGTATATCCAGCTGAACGGTCTGCTTATAGGTACAT 889
Db 11128 TTCTTTACGGTCTTTAAAAAGCGCGTATATCCAGCTGAACGGTCTGCTTATAGGTACAT 11187
Qy 890 TGAGCAACTGACTGAATTCCTCAAAATGTTCTTTACGATGCCATTCGGATATATCAACG 949
Db 11188 TGAGCAACTGACTGAATTCCTCAAAATGTTCTTTACGATGCCATTCGGATATATCAACG 11247
Qy 950 GTGGTATATCCAGTGAATTTTCTCCATTTAGCTTCCCTAGCTCCTGAAATCTCGAC 1009
Db 11248 GTGGTATATCCAGTGAATTTTCTCCATTTAGCTTCCCTAGCTCCTGAAATCTCGAT 11307
Qy 1010 AACTCAAAAAATACGCCGAGTATCTTATTTTCAATATGTTGTAAGTTCGAACTCTT 1069
Db 11308 AACTCAAAAAATACGCCGAGTATCTTATTTTCAATATGTTGTAAGTTCGAACTCTT 11367
Qy 1070 ACGTCCGATCAACGCTCTCAATTTTCGCCAAAGTTGGCCAGGCTTCCCGGTATCAACA 1129
Db 11368 ACGTCCGATCAACGCTCTCAATTTTCGCCAAAGTTGGCCAGGCTTCCCGGTATCAACA 11427
Qy 1130 GGGACACAGGATTTATTTATTCGGAAGTATCTTCGGTACAGGATTTATTTTCGGTTC 1189
Db 11428 GGGACACAGGATTTATTTATTCGGAAGTATCTTCGGTACAGGATTTATTTTCGGTTC 11487
Qy 1190 GAAGAAGATCCG 1201
Db 11488 CAAAGTGGTCG 11499

RESULT 14
AV303236
LOCUS AV303236 13515 bp DNA circular SYN 02-APR-2004
DEFINITION Shuttle vector pLEItd-KR'', complete sequence.
ACCESSION AV303236
VERSION AV303236.1 GI:32140767
KEYWORDS
SOURCE Shuttle vector pLEItd-KR''
ORGANISM Shuttle vector pLEItd-KR''
OTHER SEQUENCES; artificial sequences; vectors.
REFERENCE 1 (bases 1 to 13515)
AUTHORS Cousineau, B., Lawrence, S., Smith, D. and Belfort, M.
TITLE Retrotransposition of a bacterial group II intron
JOURNAL Nature 404 (6781), 1018-1021 (2000)
PUBMED 10801134
REFERENCE 2 (bases 1 to 13515)
AUTHORS Staddon, J. H., Bryan, E. M., Manias, D. A. and Dunny, G. M.
TITLE Conserved Target for Group II Intron Insertion in Relaxase Genes of Conjugative Elements of Gram-Positive Bacteria
JOURNAL J. Bacteriol. 186 (8), 2393-2401 (2004)
PUBMED 15060042
REFERENCE 3 (bases 1 to 13515)
AUTHORS Staddon, J. H.
TITLE Direct Submision
JOURNAL Submitted (16-MAY-2003) Microbiology, University of Minnesota, MMC 196, 420 Delaware Street S.E., Minneapolis, MN 55455, USA
FEATURES
source
1. .13515
/organism="Shuttle vector pLEItd-KR''"
/mol type="genomic DNA"
/db_xref="taxon:233895"
1623..1802
/feature="ltrBE1; partial exon"
/number=1
exon
```

```
intron
CDS
1803..5815
/feature="l1.ltrB group II intron with APHIII and td
delta(1-3)"
2376..4175
/feature="maturase"
/codon_start=1
/transl_table=11
/product="l1"
/protein_id="AAP70045.1"
/db_xref="GI:32140772"
/translation="MKPTMAILLERISKNQENIDDEVTRLYRLRLLPDYVAVONLY
SNKASNGKILDDTADGFSBEKIKKIIQSLKDGTYYPQVPMYIAKNKSKMRPLGI
PTFDDKLIQAVRILLESIVPEFEDVSHGFRPORSCHTALKTIKRFGGARWVPGD
KLANCIYHLHDLDFVLQMKRFDRESPIITPEYRELHNEIKRISHRLKJLEGEKAK
VLLEYQKRLPLTPTCTQTNKLVKYRVADDFIISVKGSKEDQWIKQQLFIHN
KLMELSEKTLTIHSSQAPFLGDIYVRRSGTIRKSGVKKTGLNGSVLLIPLQD
KIQFIPIKIAIOKDDSWFPVHRKYLIRSTDLEIITVYNSLGRGICNYVGLASNFN
QLYFAYLMYSCLTKTIASHKHGTLTKTISMPKDGSGSGWGPYIEIKQKQRRYPANFS
ECSFYQTFDEISQAPVLYGYVANTLENLAKKCCGCGTSDENTSIEIHVHNKVNKL
KGKQWMAIAKQKRTLVVCFCHRHVHHK"
4184..4195
/feature="derived from T4 phage thymidylate synthase of
GenBank Accession Number M12742"
/number=1
4196..4588
/feature="td delta(1-3) group I intron"
4589..4600
/feature="derived from T4 phage thymidylate synthase of
GenBank Accession Number M12742; partial exon"
/number=2
4845..5639
/feature="3',5',aminoglycoside phosphotransferase; kanamycin
resistance; from enterococcal plasmid pJH1 of GenBank
Accession Number V01547"
/codon_start=1
/transl_table=11
/product="APH-III"
/protein_id="AAP70046.1"
/db_xref="GI:32140773"
/translation="MAKWRISPELKLIEKYRCVKDTGMSPAKYKLVGNENLYLK
MTDSRYGTYDYVEREKDMMLEGLKLPVKYLFHFRHDCNSLLMSEADGVLSESEY
EDGSPBKILIELYACIRLFESIDISOCPTNSLDLSLAELDYLLNLDLADVCENWE
EDTPPKDRELDFLTKPESEELVSHGDLGDSNIFVKDGVSGFDLGRSGRADKW
YDIAFCVRSIREDIGERQYVELFDLLGKIPDWEKIKYIILLDEL"
5816..5906
/feature="l1rBE2; partial exon"
/number=2
complement(6756..8246)
/feature="derived from GenBank Accession Number AF041239"
/codon_start=1
/transl_table=11
/product="replication protein E"
/protein_id="AAP70043.1"
/db_xref="GI:32140770"
/translation="MNIFFVETVLDHGLLKYFKNSKIRSIITPKGSKGAIFAYS
KSMIGRVLTSSEAIQENQDTFTHTPNVRYGYADENRSYTKGSHSNLRQIN
KSTFDIDHTAKETISASDILTTAIDIGEMPTMIKSDGVQAVFVLETVPYVTSKSE
FKSVKAKIISQNIREFYFGKSLPVDLTCNHFGLIARIPRTDNVEFEDPNRYSEKEMQD
WSFKQDNKGFTRSSLTSLSGTEGKQDQEPFNLLHETKFSCEKGLIGNNVFTLC
SLAYFSSGYSIETCEYNNFNNRLDQPLEEKVIKIVRSAYSENYQANNEYITILC
KAWVSDLTSGDLFVRQGFKKKSERQVHLSWEKIDMAYISEKSDVYKPYLVY
TKKEIRVLGIPERTDLKLVKANQEIFFKIPKRGNGGQLASVSKSLLSIIKVKK
EEKESYIKALTNSFDLHETFOETLNKLAERPKTDTQDLDFSYDTG"
complement(8289..8588)
/feature="p11253 replication in gram positive bacteria;
derived from GenBank Accession Number AF041239"
/codon_start=1
/transl_table=11
/product="replication protein D"
/protein_id="AAP70044.1"
/db_xref="GI:32140771"
/translation="MNHDECKTYIKNSLLEIRKLANITYTLETFKKELEKRNLYLETKS
```


AUTHORS Partridge,S.R. and Hall,R.M.
 TITLE Complex multiple antibiotic and mercury resistance region derived from the r-det of NR1 (R100)
 JOURNAL Antimicrob. Agents Chemother. 48 (11), 4250-4255 (2004)
 PUBMED 15504849
 REFERENCE 5 (bases 4901 to 29360)
 AUTHORS Partridge,S.R. and Hall,R.M.
 TITLE Direct Submission
 JOURNAL Submitted (17-JUN-2002) CSIRO Molecular Science, PO Box 184, North Ryde, NSW 1670, Australia
 REFERENCE 6 (bases 1 to 4900)
 AUTHORS Partridge,S.R. and Hall,R.M.
 TITLE Direct Submission
 JOURNAL Submitted (25-FEB-2003) CSIRO Molecular Science, PO Box 184, North Ryde, NSW 1670, Australia
 REMARK Sequence update by submitter
 REFERENCE 7 (bases 1 to 4900)
 AUTHORS Partridge,S.R. and Hall,R.M.
 TITLE Direct Submission
 JOURNAL Submitted (12-MAR-2003) CSIRO Molecular Science, PO Box 184, North Ryde, NSW 1670, Australia
 REFERENCE 8 (bases 1 to 45325)
 AUTHORS Partridge,S.R. and Hall,R.M.
 TITLE Direct Submission
 JOURNAL Submitted (06-JAN-2004) Department of Biological Sciences, Macquarie University, Sydney, NSW 2109, Australia
 REMARK Sequence update by submitter
 COMMENT On or before Dec 20, 2004 this sequence version replaced gi:33416297, gi:33416299, gi:33416301, gi:28558819.
 FEATURES
 source
 1..45325
 /organism="Klebsiella pneumoniae"
 /mol_type="genomic DNA"
 /db_xref="taxon:573"
 /plasmid="pRMH760"
 misc_feature
 join(124..145,1473..5119)
 /note="tnp module; identical to Tn1696 tnp module in Genbank Accession Number U12338.2, with IS4321L in IRtnp"
 repeat_region
 join(124..145,1473..1488)
 /note="IRtnp of Tn1696"
 repeat_region
 146..1472
 /rpt_type=inverted
 /citation={2}
 /insertion_seq="IS4321L"
 repeat_region
 149..160
 /note="IR of IS4321"
 /rpt_type=inverted
 complement(401..1405)
 /codon_start=1
 /transl_table=11
 /product="IS4321 transposase"
 /protein_id="AAO46046.1"
 /db_xref="GI:28558821"
 translation="MENTALIGDLGNKSFHHODRRGKAVYKKFTRPKIIEPLAT
 PATTIAECCGSHFMARKEELGHPKLI.SPOFVRPFSKNKNDVDAEAI CEAS
 RPSMVFQRTESQAMALRVESLVQDKVTNNQMHAFLLFEGISVPRGAIVIS
 LTLEDNSLPLYSOLLKLQOHVYLVQI KLESOLKRLDEDEIGORULPCV
 GTLFTASTIEIGDKYASSRDEFAATGLVPROYSTGRTLLGISKRGNKIRTL
 VQCARVFIQKLEHSGKLDWVRDLLCKRSNFVTCALANKLARIANALTAQOITYVA
 "
 repeat_region
 1454..1465
 /note="IR1 of IS4321"
 /rpt_type=inverted
 complement(1484..4456)
 /gene="tnpA"
 /complement(1484..4456)
 /gene="tnpA"
 /note="Tn1696 tnpA"
 /codon_start=1
 /transl_table=11
 /product="TnpA transposase"
 /protein_id="AAO46047.1"
 /db_xref="GI:28558822"
 gene
 complement(1484..4456)
 /gene="tnpA"
 /complement(1484..4456)
 /gene="tnpA"
 /note="Tn1696 tnpA"
 /codon_start=1
 /transl_table=11
 /product="TnpA transposase"
 /protein_id="AAO46047.1"
 /db_xref="GI:28558822"

/translation="MPRRSILSATERESLLALPAKDELIRHYTFNETDLSVIRQRG
 AANRLGFAVQLCYLRFEGTGLGVDEPPFPFLLRMAAQLKMPVESWSEYSGOREOTRRE
 HLNRQTVGFQFPTMGHYRQAQVHTLALQTDKGIKVASALVENLRQSIILPANN
 AIERASABAITRANRRIYAALTDLSLLSPHROLDLKKRQDGSKVTLAWLRQSPAKP
 NSRHMLEHIELKSNQALDLPAGIEROVHONRLKIAREGGQMTPADLAKFEVORRYA
 TLVALAIEGNATVTDIELIDHRIIKLFNAAKKHQQQFQASKAINDKVMYGRIG
 QALIEAKQSGSDPPAAIEAVPWDTFAASVTEAQTLLARPADDFDLHHIGESYATLURRY
 APQFLGVLKLRAPAAKGVLDADIMLRGMNSDSARKVPADAPTAIFKPRWAKLVLTDD
 AIDTCDOYLHDLLELLEAQLATVNRMAANDLPDAIITASGLKITPLDAAVPDAQAAM
 IDOTAMLLPHLKITELMEVDEWGTFRHFTHLAKTSDAKDTLLLTLLADATNLGL
 TKMAESCPTGYAKLSWLOAHIRDETYSTALAEVNAQRPQFAGKNGDCGTTSSDG
 QNFRIGSKAESTGHINPKGSSPRTFYTHISDQIAPSAKVNVGINDSYVDLGGLL
 YHSDLRINEHYTDTAGTDFHVGMLHLLGFAPRIFAPRDLKLGTSYPRQNGLAVALRELK
 PMSDDRLNITQIRAHWEDEILATSIKQGTVTASLMLRLKIGSYPRQNGLAVALRELK
 RIERTLFTDLWLSQVELRRVHAGLNKGEARNALARAVFFYRLGEIRDRSEQQRYRA
 SGLNLVTAATVWNTVYLERATSALRGNGTALDDTLQLYLSPLGMEHINLTGDYLWRS
 SAKVGACKFRPLRPLPPA"
 complement(4459..5016)
 /gene="tnpR"
 complement(4459..5016)
 /gene="tnpR"
 /note="Tn1696 tnpR"
 /codon_start=1
 /transl_table=11
 /product="TnpR resolvase"
 /protein_id="AAM89413.2"
 /db_xref="GI:28558820"
 translation="MQGQRIGYGVSVSSFDONPEROLEGVQVAVFTDKASGDKTORPE
 LERLIATVREGDITVVVSMRNLARNLDDLRIVQGLTQGRVMEFVEGLKFTGEDSP
 MANLISVMGAFAEFERALIREROREGIVLAKORGAYRGRKKSILNSQIAELKRRVAA
 GDQKTLVARDFGISRETLYQYRED"
 misc_recomb
 5037..5058
 /note="Tn1696 resIII"
 misc_recomb
 5072..5105
 /note="Tn1696 resII"
 repeat_region
 5120..22649
 /transposon="class I integron In34"
 misc_feature
 5120..6471
 /note="5'-CS; 5'-conserved segment"
 repeat_region
 5120..5144
 /note="left-hand inverted repeat of In34; Iri"
 /rpt_type=inverted
 complement(5322..6335)
 /gene="intII"
 complement(5322..6335)
 /gene="intII"
 /codon_start=1
 /transl_table=11
 /product="IntII integrase"
 /protein_id="AAM89398.1"
 /db_xref="GI:24527229"
 translation="MKTATAPLPLRSVKVLDLQRLRIRYLHVSLRTEQAQVVMVRAF
 IIRFGVRHPATLGSSEVEAFSLWLANERKVSVSTHROALALLLFFYGVKLTCDLPWLQ
 EIGRPSRRLLPVVLTDEVRILGLFEGEHLRFALQLYTGMRISGLQURVKOLDP
 DHGTIIVREKGSNDRLMLPESLAPLSRQLSRARAWLKQDQEGSGVALPDALER
 KYPAGHSWFWFWFAQHTSDPRSGVVRHHMYDQTFQAFRAVEQAQITKPTATP
 HTLRHSATPLLRSGYDIRTVQDLLGHSVDSTTMYIHTVLKVGCGAGVSPDLAPPL
 SER"
 6220..6248
 /note="promoter Pc; transcribes cassette associated genes;
 strong version"
 6220..6225
 6243..6248
 complement(6362..6367)
 /gene="intII"
 complement(6385..6390)
 /gene="intII"
 6416..6480
 /note="attII; integron-associated recombination site
 recognized by IntII"
 6472..7062
 /note="aadB gene cassette"

[illegible]

Search completed: January 18, 2006, 04:25:16
Job time : 6465.19 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2006 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 17, 2006, 17:02:23 ; Search time 779.788 Seconds
(without alignments)
10717.687 Million cell updates/sec

Title: US-10-511-327-2
Perfect score: 1254
Sequence: 1 gatctgattgattgaacaa.....ttcgttcaatcaatcagatc 1254

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_21:.*
1: geneseqn1980s:.*
2: geneseqn1990s:.*
3: geneseqn2000s:.*
4: geneseqn2001as:.*
5: geneseqn2001bs:.*
6: geneseqn2002as:.*
7: geneseqn2002bs:.*
8: geneseqn2003as:.*
9: geneseqn2003bs:.*
10: geneseqn2003cs:.*
11: geneseqn2003ds:.*
12: geneseqn2004as:.*
13: geneseqn2004bs:.*
14: geneseqn2005s:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|-------|--------------------|
| 1 | 1254 | 100.0 | 1254 | 10 | ACF58169 Modified |
| C 2 | 1147.2 | 91.5 | 3400 | 7 | ADL18593 |
| C 3 | 1142.8 | 91.1 | 6122 | 6 | ABN84078 |
| C 4 | 1142.8 | 91.1 | 6122 | 6 | AL41112 |
| C 5 | 1142.8 | 91.1 | 6122 | 6 | AL41115 |
| C 6 | 1142.8 | 91.1 | 6125 | 6 | AL41114 |
| C 7 | 1135.2 | 90.5 | 4100 | 14 | AEA43207 |
| C 8 | 1130.8 | 90.2 | 3817 | 10 | ADD14893 |
| C 9 | 1116.4 | 89.0 | 6607 | 14 | ADZ03868 |
| C 10 | 1116.4 | 89.0 | 6608 | 14 | ADZ03870 |
| C 11 | 1116.4 | 89.0 | 9808 | 14 | ADZ68224 |
| C 12 | 1116.4 | 89.0 | 9808 | 14 | AE880052 |
| C 13 | 1116.4 | 89.0 | 9808 | 14 | AE880052 |
| C 14 | 1114.4 | 88.9 | 5838 | 2 | AAV72933 |
| C 15 | 1114.4 | 88.9 | 5838 | 2 | AAV28398 |
| C 16 | 1114.4 | 88.9 | 5838 | 2 | Az06434 |
| C 17 | 1114.4 | 88.9 | 5838 | 2 | Az222882 |
| C 18 | 1114.2 | 88.9 | 3064 | 14 | AEK87209 |
| C 19 | 1106.6 | 88.2 | 5300 | 14 | ADV09403 |
| | | | | | ADY91370 E. coli e |

| | | | | | |
|------|--------|------|-------|----|-----------|
| 20 | 1105.6 | 88.2 | 5777 | 14 | ADW46688 |
| C 21 | 1074.8 | 85.7 | 1853 | 2 | AAQ13156 |
| C 22 | 1070.4 | 85.4 | 1864 | 6 | AAQ35079 |
| C 23 | 1052.4 | 83.9 | 1906 | 10 | ADF44789 |
| C 24 | 1052.4 | 83.9 | 1992 | 10 | ADF44788 |
| C 25 | 1052.4 | 83.9 | 2191 | 10 | ADF44787 |
| C 26 | 1052.4 | 83.9 | 2714 | 10 | ADF44786 |
| C 27 | 998.2 | 79.6 | 9265 | 6 | ABK87210 |
| C 28 | 982.6 | 78.4 | 4626 | 2 | AAAT79498 |
| C 29 | 982.6 | 78.4 | 5227 | 2 | AAAT79537 |
| C 30 | 982.6 | 78.4 | 5362 | 14 | ADVA4285 |
| C 31 | 982.6 | 78.4 | 5362 | 14 | ADVA4286 |
| C 32 | 982.6 | 78.4 | 5563 | 14 | AE831043 |
| C 33 | 982.6 | 78.4 | 5587 | 14 | AE831042 |
| C 34 | 982.6 | 78.4 | 7783 | 2 | AAZ26302 |
| C 35 | 982.6 | 78.4 | 8623 | 4 | AAAF77067 |
| C 36 | 960 | 76.6 | 9460 | 5 | AAAS00466 |
| C 37 | 952.8 | 76.0 | 7507 | 2 | AAAX01427 |
| C 38 | 952.8 | 76.0 | 12495 | 12 | ADDO05573 |
| C 39 | 952.8 | 76.0 | 14468 | 8 | ACC83500 |
| C 40 | 952.8 | 76.0 | 16847 | 12 | ADO07464 |
| C 41 | 952.8 | 76.0 | 18563 | 12 | ADO07466 |
| C 42 | 952.8 | 76.0 | 18563 | 12 | ADO07465 |
| C 43 | 952.8 | 76.0 | 18565 | 12 | ADO07467 |
| C 44 | 952.8 | 76.0 | 19038 | 12 | ADO07468 |
| C 45 | 952.8 | 76.0 | 19038 | 12 | ADO07469 |

ALIGNMENTS

RESULT 1
ACF58169
ID ACF58169 standard; DNA; 1254 BP.
XX
AC ACF58169;
XX

DT 15-JAN-2004 (first entry)

DE Modified Cat-Mu(Stop)-transposon.

KW Transposon; genetic engineering; transposase; Cat-Mu; ds.

XX Synthetic.

OS Bacteriophage mu.

FN WO2003087370-A1.

PD 23-OCT-2003.

PF 14-APR-2003; 2003WO-FI000285.

PR 18-APR-2002; 2002FI-00000746.

XX (FINN-) FINNZYMES OY.

XX Savilahti H, Tieaho V;

XX WPI; 2003-845329/78.

PT New transposon nucleic acid comprising a genetically engineered translation stop signal within a transposon end sequence recognized by a transposase useful for producing deletion derivatives of polypeptide.

XX Claim 6; Page 30; Opp; English.

CC The invention relates to a transposon nucleic acid comprising a genetically engineered translation stop signal in the three reading frames at least partly within a transposon end sequence recognized by a transposase. The transposon is useful for producing deletion derivatives of polypeptide coding nucleic acids. The method involves performing a transposition reaction in the presence of a target nucleic acid containing a polypeptide coding nucleic acid and in the presence of a

CC transposon containing a genetically engineered translation stop signal
CC sequence in three reading frames at least partly within a transposon end
CC sequence recognized by a transposase; and recovering a target nucleic
CC acid having the transposon incorporated in the protein coding nucleic
CC acid. The present sequence represents a modified Cat-Mu(Stop)-transposon
XX
SQ Sequence 1254 BP; 356 A; 296 C; 264 G; 338 T; 0 U; 0 Other;

| | | | | | |
|-----------------------|-----|---|---------------|-----------|--------------|
| Query Match | | 100.0%; | Score 1254; | DB 10; | Length 1254; |
| Best Local Similarity | | 100.0%; | Pred. No. 0; | | |
| Matches 1254; | | Conservative 0; | Mismatches 0; | Indels 0; | Gaps 0; |
| QY | 1 | GATCTGATTGATTGAACGAAAAACCGGAAAGGGTTTACGATAAAATGCGAAAAACGGATCC | 60 | | |
| DB | 1 | GATCTGATTGATTGAACGAAAAACCGGAAAGGGTTTACGATAAAATGCGAAAAACGGATCC | 60 | | |
| QY | 61 | TATCGTCAATTATTACCTCCACGGGGAGAGCCTGAGCAAACTGGCCTCAGGCAATTTGAGA | 120 | | |
| DB | 61 | TATCGTCAATTATTACCTCCACGGGGAGAGCCTGAGCAAACTGGCCTCAGGCAATTTGAGA | 120 | | |
| QY | 121 | AGCACACGGTCACTGCTTCCGGTAGTCAATAAACCGGTAAACAGCAATAGACATAAG | 180 | | |
| DB | 121 | AGCACACGGTCACTGCTTCCGGTAGTCAATAAACCGGTAAACAGCAATAGACATAAG | 180 | | |
| QY | 181 | CGGCTATTTAACGACCTCGCCCTGAACCGACCGGGTCGAAATTTGCTTTTCGAAATTTCT | 240 | | |
| DB | 181 | CGGCTATTTAACGACCTCGCCCTGAACCGACCGGGTCGAAATTTGCTTTTCGAAATTTCT | 240 | | |
| QY | 241 | GCCATTCACTCGCTTATTATCACTTATTACGGCGTAGCAACAGCGGTTTAAAGGACCA | 300 | | |
| DB | 241 | GCCATTCACTCGCTTATTATCACTTATTACGGCGTAGCAACAGCGGTTTAAAGGACCA | 300 | | |
| QY | 301 | ATAACTGCTTTAAAAAATTAGCCCGCCCTGCCACTCATCGCAGTACTGTTGTAATTC | 360 | | |
| DB | 301 | ATAACTGCTTTAAAAAATTAGCCCGCCCTGCCACTCATCGCAGTACTGTTGTAATTC | 360 | | |
| QY | 361 | ATTAAGCATTTCCGCGACATGAAGCCATCAAAACCGCATCAAACTGAATCGCCAG | 420 | | |
| DB | 361 | ATTAAGCATTTCCGCGACATGAAGCCATCAAAACCGCATCAAACTGAATCGCCAG | 420 | | |
| QY | 421 | CGGCATCAGCACCTTTGTGCGCTTGCCTATAATATTGTCCTGTTGAAAAACGGGGCGAA | 480 | | |
| DB | 421 | CGGCATCAGCACCTTTGTGCGCTTGCCTATAATATTGTCCTGTTGAAAAACGGGGCGAA | 480 | | |
| QY | 481 | GNAGTTGTCATTTGGCCACGTTTAAATCAAACTGGTGAACCTACCCAGGATTCGG | 540 | | |
| DB | 481 | GNAGTTGTCATTTGGCCACGTTTAAATCAAACTGGTGAACCTACCCAGGATTCGG | 540 | | |
| QY | 541 | TGAGACGAAAAACATATTCTCAATAAACCCCTTTAGGAAATAGGCGAGTTTTCACCGTA | 600 | | |
| DB | 541 | TGAGACGAAAAACATATTCTCAATAAACCCCTTTAGGAAATAGGCGAGTTTTCACCGTA | 600 | | |
| QY | 601 | ACACGCCACATCTTCGGAATATATGTAGAAACTGCCGGAATCGTCGTGTAATTCAT | 660 | | |
| DB | 601 | ACACGCCACATCTTCGGAATATATGTAGAAACTGCCGGAATCGTCGTGTAATTCAT | 660 | | |
| QY | 661 | CCAGAGCGATGAAGAGTTTCACTTGTCTCATGGAACCGGTGAACAGGGTGAACACT | 720 | | |
| DB | 661 | CCAGAGCGATGAAGAGTTTCACTTGTCTCATGGAACCGGTGAACAGGGTGAACACT | 720 | | |
| QY | 721 | ATCCCATATCACAGCTACCGCTCTTTTCATTCGCCATAGCTAAATCCGGATGAGCATTCAT | 780 | | |
| DB | 721 | ATCCCATATCACAGCTACCGCTCTTTTCATTCGCCATAGCTAAATCCGGATGAGCATTCAT | 780 | | |
| QY | 781 | CAGGCGGGCAGAATGTGAATAAAGCCGGATAAAACTTGTCTTATTTTCTTTACGGT | 840 | | |
| DB | 781 | CAGGCGGGCAGAATGTGAATAAAGCCGGATAAAACTTGTCTTATTTTCTTTACGGT | 840 | | |
| QY | 841 | CTTTAAAAAGCCGTAATATCAGCTGAACGGTCTGGTTATAGGTACATTGAGCACTGA | 900 | | |
| DB | 841 | CTTTAAAAAGCCGTAATATCAGCTGAACGGTCTGGTTATAGGTACATTGAGCACTGA | 900 | | |
| QY | 901 | CTGAAATGCCTCAAAATGTTCTTTTACGATGCCATTGGGATATATCAACGGTGTATATCC | 960 | | |

| | | | | |
|------------|-------------|---|------|--|
| Db | 901 | CTGAAATGCCTCAAAATGTTCTTTTACGATGCCATTGGGATATATCAACGGTGTATATCC | 960 | |
| Qy | 961 | AGTGATTTTTCCTCCATTTTAGCTTCTTAGCTCTCGAAATCTCGACAACCTCAAAAAA | 1020 | |
| Db | 961 | AGTGATTTTTCCTCCATTTTAGCTTCTTAGCTCTCGAAATCTCGACAACCTCAAAAAA | 1020 | |
| Qy | 1021 | TAGCCCGGTAGTGCATCTTATTTTCATTTATGTTGAAAGTTGGAACCTCTTACGTGCCGATC | 1080 | |
| Db | 1021 | TAGCCCGGTAGTGCATCTTATTTTCATTTATGTTGAAAGTTGGAACCTCTTACGTGCCGATC | 1080 | |
| Qy | 1081 | AACGCTCTCATTTTCGCCCAAAAGTTGGCCACAGGCTTCCCGGTATCAACAGGACACCAAG | 1140 | |
| Db | 1081 | AACGCTCTCATTTTCGCCCAAAAGTTGGCCACAGGCTTCCCGGTATCAACAGGACACCAAG | 1140 | |
| Qy | 1141 | ATTTATTTATTTCTCGGAAGTGCATCTTCGCTCACAGGTATTTTATTCGGTCTGAAAAGGATCC | 1200 | |
| Db | 1141 | ATTTATTTATTTCTCGGAAGTGCATCTTCGCTCACAGGTATTTTATTCGGTCTGAAAAGGATCC | 1200 | |
| Qy | 1201 | GTTTTTCGCAATTATCGTGAAACGGCTTTTCGCGTTTTTCGTTCAATCAATCAGATC | 1254 | |
| Db | 1201 | GTTTTTCGCAATTATCGTGAAACGGCTTTTCGCGTTTTTCGTTCAATCAATCAGATC | 1254 | |
| RESULT 2 | | | | |
| ADL18593/C | | | | |
| ID | ADL18593 | standard; DNA; 3400 BP. | | |
| XX | ADL18593; | | | |
| AC | | | | |
| DT | 06-MAY-2004 | (first entry) | | |
| XX | | Plasmid pBC SK+ nucleotide sequence SEQ ID NO:1. | | |
| DE | | gene transfer; product vector; purification; detection; plasmid; pBC SK+; | | |
| KW | | circular; gene; ds. | | |
| XX | | Synthetic. | | |
| OS | | | | |
| PN | | WO200268670-A1. | | |
| XX | | | | |
| PD | | 06-SEP-2002. | | |
| XX | | | | |
| PF | | 15-FEB-2002; 2002WO-US004454. | | |
| XX | | | | |
| PR | | 26-FEB-2001; 2001US-00793372. | | |
| XX | | (STRA-) STRATAGENE. | | |
| PA | | | | |
| XX | | Carstens C; | | |
| XX | | | | |
| PI | | WPI; 2002-707008/76. | | |
| XX | | | | |
| DR | | | | |
| XX | | | | |
| PT | | Transferring a gene to a product vector, useful for generating | | |
| PT | | recombinant vectors for expression in cell or host, comprises contacting | | |
| PT | | in vitro a first and second vector and introducing the co-integrate | | |
| PT | | vector into a prokaryotic cell. | | |
| XX | | | | |
| PS | | Example 1; SEQ ID NO 1; 50pp; English. | | |
| XX | | | | |
| CC | | The present invention describes a method for transferring genes to a | | |
| CC | | product vector comprising contacting in vitro a first and second vector | | |
| CC | | and introducing the co-integrate vector into a prokaryotic cell to permit | | |
| CC | | the formation of a product vector. Also described: (1) a pair of vectors | | |
| CC | | (i) comprising: (a) a first vector comprising a gene or a cloning site | | |
| CC | | for the insertion of a gene, a gene encoding a first selectable marker, a | | |
| CC | | double-stranded origin of replication and a site-specific recombination | | |
| CC | | recognition site, where the gene is interposed between the double- | | |
| CC | | stranded origin of replication of a rolling circle replicon and the site- | | |
| CC | | specific recombination recognition site; and (b) a second vector | | |
| CC | | comprising a negative selectable marker, a double-stranded and double- | | |
| CC | | stranded origin of replication of a rolling circle replicon, a site- | | |
| CC | | specific recombination recognition site and a gene encoding a second | | |

selectable marker, where the gene encoding the negative selectable marker is interposed between the double-stranded origin of replication of a rolling circle replicon and the site-specific recognition site, where one or both of the vectors have no second site-specific recombinase recognition site between the double-stranded origin of replication and the site-specific recombinase recognition site; (2) a product vector comprising a gene, a double-stranded origin of replication of a rolling circle replicon, a site-specific recombinase site, a single-stranded origin of replication and a nucleic acid sequence encoding a second selectable marker, where the gene is interposed between the double-stranded origin of replication of a rolling circle replicon and the site-specific recombinase recognition site; and (3) a kit for the transfer of gene to a product vector comprising (i) and packaging materials. The method is useful for generating recombinant vectors. These recombinant vectors are useful in expressing mammalian cell and bacterial hosts, purification of the native protein by employing specialised purification tags and CC detection of interaction with other proteins. The present sequence CC represents the nucleotide sequence of plasmid pBC SK+, which is used in CC the exemplification of the present invention.

Sequence 3400 BP; 877 A; 804 C; 863 G; 856 T; 0 U; 0 Other;

Query Match 91.5%; Score 1147.2; DB 7; Length 3400;

Best Local Similarity 98.5%; Pred. No. 0; Mismatches 18; Indels 0; Gaps 0; Matches 1158; Conservative 0;

QY 50 AAAACGGATCCTATCGTCAATATTACCTCCACGGGAGAGCTCAGCAAACTGCGCTCA 109
 DB 3053 AAAAAGGATCATATCGTCAATATTACCTCCACGGGAGAGCTCAGCAAACTGCGCTCA 2994
 QY 110 GGCATTTCGAGAACGACGCGTCACACTGCTTCGGTAGTCAATAAACCGGTAAACGACA 169
 DB 2993 GGCATTTCGAGAACGACGCGTCACACTGCTTCGGTAGTCAATAAACCGGTAAACGACA 2934
 QY 170 ATAGACATAAGCGGCTATTAAAGACCTGCGCTGAAACGAGACCGGGTCAATTCGT 229
 DB 2933 ATAGACATAAGCGGCTATTAAAGACCTGCGCTGAAACGAGACCGGGTCAATTCGT 2874
 QY 230 TTCGAATTTCTGCCATTTCATCGCTTATTATCAGCTTATTCAGGCGTAGCAACAGCGCTT 289
 DB 2873 TTCGAATTTCTGCCATTTCATCGCTTATTATCAGCTTATTCAGGCGTAGCAACAGCGCTT 2814
 QY 290 TAAGGGACCAATAACTGCTTTAAAAAAATTAACGCGCGCGCTGCCACTCATCGAGTAC 349
 DB 2813 TAAGGGACCAATAACTGCTTTAAAAAAATTAACGCGCGCGCTGCCACTCATCGAGTAC 2754
 QY 350 TGTGTAAATTCATTAAAGCATTCGCGGACATGGAAGCCATCAAAACGCGCATGTAAC 409
 DB 2753 TGTGTAAATTCATTAAAGCATTCGCGGACATGGAAGCCATCAAAACGCGCATGTAAC 2694
 QY 410 TGAATCGCGCAGCGCATCAGCACCTTGTGCGCTTGTGATATATTTCCCATGTTGAAA 469
 DB 2693 TGAATCGCGCAGCGCATCAGCACCTTGTGCGCTTGTGATATATTTCCCATGTTGAAA 2634
 QY 470 ACGGGGGGGAAGAAGTTGTTCATATGGCCAGCTTTAAATCAAACTGGTGAATCTCA 529
 DB 2633 ACGGGGGGGAAGAAGTTGTTCATATGGCCAGCTTTAAATCAAACTGGTGAATCTCA 2574
 QY 530 CAGGATTTGGCTGAGACGAAAAACATATTCATATAAACCTTTTAGGAAATAGCCAGG 589
 DB 2573 CAGGATTTGGCTGAGACGAAAAACATATTCATATAAACCTTTTAGGAAATAGCCAGG 2514
 QY 590 TTTTCACCGTAACAGCCACATCTTGCATATATGTAGAACTGCGGAATCGTCG 649
 DB 2513 TTTTCACCGTAACAGCCACATCTTGCATATATGTAGAACTGCGGAATCGTCG 2454
 QY 650 TGGTATTTCATCTCAGAGCGATGAAAAAGTTTCAGTTTGTCTATGAAAAAGCGTGAACAA 709
 DB 2453 TGGTATTTCATCTCAGAGCGATGAAAAAGTTTCAGTTTGTCTATGAAAAAGCGTGAACAA 2394
 QY 710 GGGTGAACATATCCCATATCACCAGCTCACCGTCTTTTCATTCGCATACGTAATTCGGA 769
 DB 2393 GGGTGAACATATCCCATATCACCAGCTCACCGTCTTTTCATTCGCATACGTAATTCGGA 2334

QY 770 TGAGCATTTCATCAGCGCGGCAAGAAATGTGAATAAAGCCGCGATAAACTTGTGCTTATTT 829
 DB 2333 TGAGCATTTCATCAGCGCGGCAAGAAATGTGAATAAAGCCGCGATAAACTTGTGCTTATTT 2274
 QY 830 TTCTTTACGGTCTTTAAAAAGCGCGTATATCCAGCTGAACGGTCTGTTATAGGTACAT 889
 DB 2273 TTCTTTACGGTCTTTAAAAAGCGCGTATATCCAGCTGAACGGTCTGTTATAGGTACAT 2214
 QY 890 TGAGCAACTGACTCAAAATGCTTCAAAATGTTCTTTACGATGCCATGGGATATATCAACG 949
 DB 2213 TGAGCAACTGACTCAAAATGCTTCAAAATGTTCTTTACGATGCCATGGGATATATCAACG 2154
 QY 950 GTGGTATATCAGTGAATTTTTTCTCCATTTTAGCTTCTTTCAGTCTCTGAAAAATCTCGAC 1009
 DB 2153 GTGGTATATCAGTGAATTTTTTCTCCATTTTAGCTTCTTTCAGTCTCTGAAAAATCTCGAC 2094
 QY 1010 AACTCAAAAATACGCCCGGTAGTGAATTTTATTCATATGTTGAAAGTTGGAACCTCTT 1069
 DB 2093 AACTCAAAAATACGCCCGGTAGTGAATTTTATTCATATGTTGAAAGTTGGAACCTCTT 2034
 QY 1070 ACGTCCGATCAACGCTCTCATTTTCGCCAAAGTTGGCCAGGGCTTCCCGGTATCAACA 1129
 DB 2033 ACGTCCGATCAACGCTCTCATTTTCGCCAAAGTTGGCCAGGGCTTCCCGGTATCAACA 1974
 QY 1130 GGGACACAGGATTTATTTATTCGCAAGTGTATCTTCGTCACAGGTATTTATTCGGTTC 1189
 DB 1973 GGGACACAGGATTTATTTATTCGCAAGTGTATCTTCGTCACAGGTATTTATTCGGTTC 1914
 QY 1190 GAAAGGATCGTTTTCGCAATTTATCGTGAACCGT 1225
 DB 1913 GAAAGGATCGTTTTCGCAAGTGTATCTTTTGTGATAATCT 1878

RESULT 3
 ABN84078/c
 ID ABN84078 standard; DNA; 6122 BP.
 XX AC ABN84078;
 XX DT 23-SEP-2002 (first entry)
 XX DE Plasmid pRL5-CAT.
 XX KW Plasmid pRL5-CAT; vector; antibody; gene; ds.
 XX OS Unidentified.
 XX FH Key Location/Qualifiers
 XX CDS 591..800
 FT /tag= a
 FT /partial
 FT /product= "Amp frag"
 FT /note= "the CDS does not include a stop codon"
 FT CDS 1013..1672
 FT /tag= b
 FT /product= "Chloramphenicol transferase"
 FT misc_feature complement (2052..2651)
 FT /tag= c
 FT /note= "ori"
 FT promoter 2831..3046
 FT /tag= d
 FT /note= "lac promoter"
 FT misc_signal 3009..3029
 FT /tag= e
 FT /note= "lac rep site"
 FT RBS 3036..3039
 FT /tag= f
 FT CDS 3050..3115
 FT /tag= g
 FT /partial
 FT /product= "OmpA leader"
 FT /note= "the CDS does not include a stop codon"


```

QY 1190 GAAAG 1195
Db 801 GAAAG 796

RESULT 4
AAL41112/c
ID AAL41112 standard; DNA; 6122 BP.
XX
AC AAL41112;
XX
DT 16-OCT-2002 (first entry)
XX
DE Plasmid pRL5 CAT nucleic acid sequence.
XX
KW Plasmid; vector; replication; expression; foreign genetic; bacteria;
KW cyanobacterium; Escherichia coli; pComb3X plasmid; pRL5; pRL5 asc-CAT;
KW pRL5-CAT; pRL5 bsi-CAT; gene; ds.
XX
OS Unidentified.
XX
FH Key Location/Qualifiers
FT CDS 1013..1672
FT /tag= a
FT /product= "CAT amino acid sequence"
FT CDS 3050..3115
FT /tag= b
FT /product= "Omp A leader amino acid sequence"
FT CDS 3953..4267
FT /tag= c
FT /product= "Kappa constant region amino acid sequence"
FT CDS 4298..4363
FT /tag= d
FT /product= "pel B leader amino acid sequence"
FT CDS 5179..5690
FT /tag= e
FT /product= "CHI, His6 tag, HA tag, and gene III amino acid
sequence"
XX
PN WO200246436-A2.
XX
PD 13-JUN-2002.
XX
XX
XX 07-DEC-2001; 2001WO-US046516.
XX
XX 08-DEC-2000; 2000US-0254411P.
XX
XX (ALEX-) ALEXION PHARM INC.
XX
XX Bowdish KS, Barbas-Frederickson S, Wild M, McWhirter J;
XX
XX WPI; 2002-537570/57.
XX
XX P-PSDB; AAO22535, AAO22536, AAO22537, AAO22538, AAO22539.
XX
XX Novel plasmid useful in cloning and expression of foreign genetic
information.
XX
XX Claim 1; Fig 7A-E; 39pp; English.
XX
XX The invention relates to a plasmid with a nucleic acid sequence of 6122,
CC 6125, or 5683 base pairs fully defined in the specification. The
CC invention more specifically relates to novel vectors capable of
CC replication and expression of foreign genetic information in bacteria,
CC such as, for example, cyanobacterium and E. coli. The new vectors have
CC been designed to overcome certain drawbacks of the pComb3X plasmid. These
CC new vectors include pRL5, pRL5-CAT, pRL5 asc-CAT and pRL5 bsi-CAT. This
CC polynucleotide sequence represents the plasmid pRL5-CAT nucleic acid
XX sequence of the invention
XX
XX Sequence 6122 BP; 1449 A; 1571 C; 1639 G; 1463 T; 0 U; 0 Other;
SQ 91.1%; Score 1142.8; DB 6; Length 6122;
Query Match

```

```

Best Local Similarity 99.8%; Pred. No. 0;
Matches 1144; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 50 AAAACGGATCTATCGTCAATTTATCTCTCCACGGGAGAGCCTTGAGCAAACTGSCCTCA 109
Db |||||
1941 AAAAAGGATCATATCGTCAATTTATCTCTCACGGGAGAGCCTTGAGCAAACTGSCCTCA 1882
QY 110 GGCATTTTGAGAACACACGGTCACTGCTTCCGGTAGTCAATAAACCGGTAAACACGCA 169
Db |||||
1881 GGCATTTTGAGAACACACGGTCACTGCTTCCGGTAGTCAATAAACCGGTAAACACGCA 1822
QY 170 ATAGACATPACGGGCTATTTAAACGACCTGCGCTGAACCGAGACCGGTCGAATTTGCT 229
Db |||||
1821 ATAGACATPACGGGCTATTTAAACGACCTGCGCTGAACCGAGACCGGTCGAATTTGCT 1762
QY 230 TTCCGAATTTTCGCATTCATCCGCTTATTATCACTTTTTCAGGGGTAGCAACAGGCGTT 289
Db |||||
1761 TTCCGAATTTTCGCATTCATCCGCTTATTATCACTTTTTCAGGGGTAGCAACAGGCGTT 1702
QY 290 TAAGGGCACCAATAACTGCTTTAAAAAATTAACCCCGCCCTGCCACTCATCGCAGTAC 349
Db |||||
1701 TAAGGGCACCAATAACTGCTTTAAAAAATTAACCCCGCCCTGCCACTCATCGCAGTAC 1642
QY 350 TGTGTGTAATTCATTAAGCATTTCTCCGACATGGAAGCATCAAAAACGGCATGATGAACC 409
Db |||||
1641 TGTGTGTAATTCATTAAGCATTTCTCCGACATGGAAGCATCAAAAACGGCATGATGAACC 1582
QY 410 TGAATCCCGAGCGGCATCAGCACCTTGTGCGCTTTCGCTATATATTTGCCCATCGTGAAA 469
Db |||||
1581 TGAATCCCGAGCGGCATCAGCACCTTGTGCGCTTTCGCTATATATTTGCCCATCGTGAAA 1522
QY 470 ACGGGGCGGAAGAGTTGTCCATATTCGCCACGCTTTAAATCAAAACTGGTGAACACTCAC 529
Db |||||
1521 ACGGGGCGGAAGAGTTGTCCATATTCGCCACGCTTTAAATCAAAACTGGTGAACACTCAC 1462
QY 530 CAGGGATGGCTGAGACGAAACATATTTCTCAATAAACCTTTAGGGAATAGGCCAGG 589
Db |||||
1461 CAGGGATGGCTGAGACGAAACATATTTCTCAATAAACCTTTAGGGAATAGGCCAGG 1402
QY 590 TTTTCACCGTAACACGCCACATCTTGGCAATATATGTGAAACTCCCGGAATCTGTCG 649
Db |||||
1401 TTTTCACCGTAACACGCCACATCTTGGCAATATATGTGAAACTCCCGGAATCTGTCG 1342
QY 650 TGTATTTCACTCCAGAGCGATGAAACGTTTCAGTTTGTCTCATGGAACCGTGTAAACA 709
Db |||||
1341 TGTATTTCACTCCAGAGCGATGAAACGTTTCAGTTTGTCTCATGGAACCGTGTAAACA 1282
QY 710 GGGTGAACACTATCCCATATCACAGCTCACCGTCTTTTCATTTGCCATACGTAATCCGGA 769
Db |||||
1281 GGGTGAACACTATCCCATATCACAGCTCACCGTCTTTTCATTTGCCATACGTAATCCGGA 1222
QY 770 TGACATTCATCAGCGCGCAAGNATGTGAATTAAGCGCGGATAAACTTGTGCTTATTT 829
Db |||||
1221 TGACATTCATCAGCGCGCGCAAGNATGTGAATTAAGCGCGGATAAACTTGTGCTTATTT 1162
QY 830 TTCTTTACGGTCTTTAAAAAGGCGTAATATCCAGCTGAACCGTCTGCTTATAGGTACAT 889
Db |||||
1161 TTCTTTACGGTCTTTAAAAAGGCGTAATATCCAGCTGAACCGTCTGCTTATAGGTACAT 1102
QY 890 TGACAACTGACTGAAATGCTCAAAATGTTCTTTACGATGCCATTCGGGATATATCAACG 949
Db |||||
1101 TGACAACTGACTGAAATGCTCAAAATGTTCTTTACGATGCCATTCGGGATATATCAACG 1042
QY 950 GTGTATATCCAGTGATTTTCTCCATTTAGCTTCTTCTAGCTCCCTGAAAATCTCGAC 1009
Db |||||
1041 GTGTATATCCAGTGATTTTCTCCATTTAGCTTCTTCTAGCTCCCTGAAAATCTCGAC 982
QY 1010 AACTCAAAAATACGCCCGGTAGTATCTTATTTATTTATTTATTTATTTATTTATTTATTT 1069
Db |||||
981 AACTCAAAAATACGCCCGGTAGTATCTTATTTATTTATTTATTTATTTATTTATTTATTT 922
QY 1070 ACGTGGCGGATCAACGCTCTCATTTTCGCCAAAAGTTGCCCGGCGCTTCCCGGTATCAACA 1129
Db |||||

```

Db 921 ACGTCCGATCAACGCTCTCATTTTGGCCAAAAGTTGGCCACGGGCTTCCCGGTATCAACA 862

Qy 1130 GGGACACAGATTTATTTATTTCTGCGAAGTATCTTCGTCACAGGATTTATTTTCGGTC 1189

Db 861 GGGACACAGATTTATTTATTTCTGCGAAGTATCTTCGTCACAGGATTTATTTTCGGTC 802

Qy 1190 GAAAAG 1195

Db 801 GAAAAG 796

RESULT 5

AAAL41115/c

ID AAAL41115 standard; DNA; 6122 BP.

XX

AC AAAL41115;

XX

DT 16-OCT-2002 (first entry)

XX

DE Plasmid pRL5 bsi-CAT nucleic acid sequence.

XX

KW Plasmid; vector; replication; expression; foreign genetic; bacteria;

KW cyanobacterium; Escherichia coli; pComb3x plasmid; pRL5; pRL5 asc-CAT;

KW pRL5-CAT; pRL5 bsi-CAT; gene; ds.

XX

OS Unidentified.

XX

FH Key

FT CDS Location/Qualifiers

FT 1013..1672

FT /tag= a

FT /product= "CAT amino acid sequence"

FT CDS 3050..3115

FT /tag= b

FT /product= "Omp A leader amino acid sequence"

FT CDS 3947..4267

FT /tag= c

FT /product= "Kappa constant region amino acid sequence"

FT CDS 4298..4363

FT /tag= d

FT /product= "pel B leader amino acid sequence"

FT CDS 5179..6090

FT /tag= e

FT /product= "CHI, His6 tag, HA tag, and gene III amino acid sequence"

XX

WO200246436-A2.

PN

XX

PD 13-JUN-2002.

XX

PF 07-DEC-2001; 2001WO-US046516.

XX

PR 08-DEC-2000; 2000US-0254411P.

XX

PA (ALEX-) ALEXION PHARM INC.

XX

PI Bowdish KS, Barbas-Frederickson S, Wild M, McWhirter J;

XX

DR WPI; 2002-537570/57.

XX

DR P-PSDB; AAO22535, AAO22536, AAO22538, AAO22539, AAO22541.

XX

PT Novel plasmid useful in cloning and expression of foreign genetic

PT information.

XX

PS Disclosure; Fig 11A-E; 39pp; English.

XX

CC The invention relates to a plasmid with a nucleic acid sequence of 6122,

CC 6125, or 5683 base pairs fully defined in the specification. The

CC invention more specifically relates to novel vectors capable of

CC replication and expression of foreign genetic information in bacteria,

CC such as, for example, cyanobacterium and E. coli. The new vectors have

CC been designed to overcome certain drawbacks of the pComb3x plasmid. These

CC new vectors include pRL5, pRL5-CAT, pRL5 asc-CAT and pRL5 bsi-CAT. This

CC polynucleotide sequence represents the plasmid pRL5 bsi-CAT nucleic acid

CC sequence of the invention

XX

SQ Sequence 6122 BP; 1447 A; 1571 C; 1641 G; 1463 T; 0 U; 0 Other;

Query Match 91.1%; Score 1142.8; DB 6; Length 6122;

Best Local Similarity 99.8%; Pred. No. 0;

Matches 1144; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 50 AAAACGGATCCTATCGTCAATTATTACCTCCAGGGGAGAGCCTGAGCAAACTGGCCTCA 109

Db 1941 AAAAAGGATCATATCGTCAATTATTACCTCCAGGGGAGAGCCTGAGCAAACTGGCCTCA 1882

Qy 110 GGCATTTGAGAAGCACACGGGTACACATGCTTCCGGTAGTCAATAAACCGGTAAACACGACA 169

Db 1881 GGCATTTGAGAAGCACACGGGTACACATGCTTCCGGTAGTCAATAAACCGGTAAACACGACA 1822

Qy 170 ATAGACATAAGCGGCTATTTTAAACGACCCCTGGAACCGACGACCGGGTCAAAATTTGCT 229

Db 1821 ATAGACATAAGCGGCTATTTTAAACGACCCCTGGAACCGACGACCGGGTCAAAATTTGCT 1762

Qy 230 TTGCAATTTCTGCCATTCATCCGCTTATTAATCACTTATTCAGGCGTAGCAACACGAGCGTT 289

Db 1761 TTGCAATTTCTGCCATTCATCCGCTTATTAATCACTTATTCAGGCGTAGCAACACGAGCGTT 1702

Qy 290 TAAGGCACCAATAAATGCTTAAAAAATTTAGCCGCCGCCCTGCCACTCATCGCAGTAC 349

Db 1701 TAAGGCACCAATAAATGCTTAAAAAATTTAGCCGCCGCCCTGCCACTCATCGCAGTAC 1642

Qy 350 TGTGTAAATTCATTAAGCATTTCTGCCGACATGGAAGCCATCAAAACGGCATGATGAACC 409

Db 1641 TGTGTAAATTCATTAAGCATTTCTGCCGACATGGAAGCCATCAAAACGGCATGATGAACC 1582

Qy 410 TGAATCCGACGGCGCATCAGCACCTTCTGCGTATAATAATTTGCCCATCGTGAAG 469

Db 1581 TGAATCCGACGGCGCATCAGCACCTTCTGCGTATAATAATTTGCCCATCGTGAAG 1522

Qy 470 ACGGGGCGAAGAGTTGTCCATATTGGCCACAGTTTAAATCAAACTGGTGAACCTCAC 529

Db 1521 ACGGGGCGAAGAGTTGTCCATATTGGCCACAGTTTAAATCAAACTGGTGAACCTCAC 1462

Qy 530 CAGGATTTGGCTGAGACGAAAAACATATTTCTCAATAAACCCCTTTAGGGAATAAGCCAGG 589

Db 1461 CAGGATTTGGCTGAGACGAAAAACATATTTCTCAATAAACCCCTTTAGGGAATAAGCCAGG 1402

Qy 590 TTTTCACCGTAACACGCCCATCTTGGCAATATATGTAGAACTCCCGGAATTCGTCG 649

Db 1401 TTTTCACCGTAACACGCCCATCTTGGCAATATATGTAGAACTCCCGGAATTCGTCG 1342

Qy 650 TGTATTTCACCTCCAGAGCGATGAAAAAGTTTTCAGTTTGGCTCATGAAAAACGGTGAACAA 709

Db 1341 TGTATTTCACCTCCAGAGCGATGAAAAAGTTTTCAGTTTGGCTCATGAAAAACGGTGAACAA 1282

Qy 710 GGGTGAACACTATCCCATATCAACAGCTCACCGTCTTTTCATTTGCCATACGTAATTCGGA 769

Db 1281 GGGTGAACACTATCCCATATCAACAGCTCACCGTCTTTTCATTTGCCATACGTAATTCGGA 1222

Qy 770 TGAGCATTCATCAGGCGGCAAGATCTGAATAAGCGCGATAAACTTGTGCTTATTT 829

Db 1221 TGAGCATTCATCAGGCGGCAAGATCTGAATAAGCGCGATAAACTTGTGCTTATTT 1162

Qy 830 TTCTTTTACGGTCTTTTAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACAT 889

Db 1161 TTCTTTTACGGTCTTTTAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACAT 1102

Qy 890 TGAGCAACTGACTGAAATGCTTAAAAATGTTCTTTAGTGCCTATTTGGATATATCAACG 949

Db 1101 TGAGCAACTGACTGAAATGCTTAAAAATGTTCTTTAGTGCCTATTTGGATATATCAACG 1042

Qy 950 GTGCTATATCCAGTGATTTTCTCCATTTTACGTTTCTTGTAGTCTCTGAAAAATCTCGAC 1009

Db 1041 GTGCTATATCCAGTGATTTTCTCCATTTTACGTTTCTTGTAGTCTCTGAAAAATCTCGAC 982

Qy 1010 AACTCAAAAAAATACGCCCGGTAGTGATCTTTATTTTATTATGGTGAAAAAGTTGGAACTCTT 1069

Db 981 AACTCAAAATACCCCGGTAGTATCTATTTTCATTATGTTGAAGTTGGAACCTCTT 922
 Qy 1070 ACGTCCGATCAACGCTCTCATTTTCGCCAAAGTTGGCCAGGCTTCCCGGTATCAACA 1129
 Db 921 ACGTCCGATCAACGCTCTCATTTTCGCCAAAGTTGGCCAGGCTTCCCGGTATCAACA 862
 Qy 1130 GGGACACAGGATTTATTTATTTCTCGGAAGTGAATCTTCGTCACAGGATTTATTCGTC 1189
 Db 861 GGGACACAGGATTTATTTATTTCTCGGAAGTGAATCTTCGTCACAGGATTTATTCGTC 802
 Qy 1190 GAAAG 1195
 Db 801 GAAAG 796

RESULT 6

AAL41114/c
 ID AAL41114 standard; DNA; 6125 BP.
 XX AAL41114;
 XX AC
 XX DT 16-OCT-2002 (first entry)
 XX DE Plasmid pRL5 CAT-Asc nucleic acid sequence.
 XX KW Plasmid; vector; replication; expression; foreign genetic; bacteria;
 KW cyanobacterium; Escherichia coli; pComb3X plasmid; pRL5; pRL5 asc-CAT;
 KW pRL5-CAT; pRL5 bai-CAT; gene; ds.
 XX OS Unidentified.
 XX FH Location/Qualifiers
 FT CDS 1013..1672
 FT /tag= a
 FT /product= "CAT amino acid sequence"
 FT CDS 3050..3115
 FT /tag= b
 FT /product= "Omp A leader amino acid sequence"
 FT CDS 3953..4270
 FT /tag= c
 FT /product= "Kappa constant region amino acid sequence"
 FT CDS 4301..4366
 FT /tag= d
 FT /product= "pel B leader amino acid sequence"
 FT CDS 5182..6093
 FT /tag= e
 FT /product= "CH1, His6 tag, HA tag, and gene III amino acid sequence"

WO200246436-A2.

13-JUN-2002.

07-DEC-2001; 2001WO-US046516.

08-DEC-2000; 2000US-0254411P.

(ALEX-) ALEXION PHARM INC.

Bowdich KS, Barbas-Frederickson S, Wild M, McWhirter J;

WPI; 2002-537570/57.

P-PSDB; AAO22535, AAO22536, AAO22538, AAO22539, AAO22540.

Novel plasmid useful in cloning and expression of foreign genetic information.

Claim 1; Fig 9A-E; 39pp; English.

The invention relates to a plasmid with a nucleic acid sequence of 6122, 6125, or 5683 base pairs fully defined in the specification. The invention more specifically relates to novel vectors capable of

CC replication and expression of foreign genetic information in bacteria, such as, for example, cyanobacterium and E. coli. The new vectors have been designed to overcome certain drawbacks of the pComb3X plasmid. These new vectors include pRL5, pRL5-CAT, pRL5 asc-CAT and pRL5 bai-CAT. This CC polynucleotide sequence represents the plasmid pRL5 CAT-Asc nucleic acid sequence of the invention
 XX
 SQ Sequence 6125 BP; 1447 A; 1573 C; 1644 G; 1461 T; 0 U; 0 Other;
 Query Match 91.1%; Score 1142.8; DB 6; Length 6125;
 Best Local Similarity 99.8%; Pred. No. 0;
 Matches 1144; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 Qy 50 AAAACGGATCTATCGTCAATTTATCTCCACGGGAGAGCCTGAGCAAACTGCGCTCA 109
 Db 1941 AAAAAGGATCATATCGTCAATTTATCTCCACGGGAGAGCCTGAGCAAACTGCGCTCA 1882
 Qy 110 GGCATTTGGAAGCACAGGTCACACTGCTTCCGCTAGTCAATTAACCGGTAAACACGCA 169
 Db 1881 GGCATTTGGAAGCACAGGTCACACTGCTTCCGCTAGTCAATTAACCGGTAAACACGCA 1822
 Qy 170 ATAGACATAAGCGGCTATTTAAACGACCTGCGCTTGAACCGGAGCAGCCGGTTCGTT 229
 Db 1821 ATAGACATAAGCGGCTATTTAAACGACCTGCGCTTGAACCGGAGCAGCCGGTTCGTT 1762
 Qy 230 TTCGAATTTTCGCCATTCATCCGCTTATTAATCACTTATTCAGGGGTAGCAACCGCGTT 289
 Db 1761 TTCGAATTTTCGCCATTCATCCGCTTATTAATCACTTATTCAGGGGTAGCAACCGCGTT 1702
 Qy 290 TAAGGGACCAATTAACCTGCTTAAAAAATTAAGCCCGCCCTGCGCACTCATCGCAGTAC 349
 Db 1701 TAAGGGACCAATTAACCTGCTTAAAAAATTAAGCCCGCCCTGCGCACTCATCGCAGTAC 1642
 Qy 350 TGTGTAATTCATTAAGCATTTCTCGACATGGAAGCCATCAAAACGGCATGATGAACC 409
 Db 1641 TGTGTAATTCATTAAGCATTTCTCGACATGGAAGCCATCAAAACGGCATGATGAACC 1582
 Qy 410 TGAATTCGCGCAGCGGCATCAGCACCTTTGTCGCTTTCGCTATAATATTTGCCCATGTTGAAA 469
 Db 1581 TGAATTCGCGCAGCGGCATCAGCACCTTTGTCGCTTTCGCTATAATATTTGCCCATGTTGAAA 1522
 Qy 470 ACGGGGGGGAAGAAGTTGTCCATATGTCACAGCTTTAAATCAAAACTGGTGAACCTACC 529
 Db 1521 ACGGGGGGGAAGAAGTTGTCCATATTTGCCACGTTTAAATCAAAACTGGTGAACCTACC 1462
 Qy 530 CAGGGATTTGGCTGAGACGGAACACATATTCCTCAATAAACCCCTTTAGGGAATAGGCCAGG 589
 Db 1461 CAGGGATTTGGCTGAGACGGAACACATATTCCTCAATAAACCCCTTTAGGGAATAGGCCAGG 1402
 Qy 590 TTTTCACCGTAACACGCGCACATCTTTCGGAATATATGTAGAAACTGCGGAAATCGTCG 649
 Db 1401 TTTTCACCGTAACACGCGCACATCTTTCGGAATATATGTAGAAACTGCGGAAATCGTCG 1342
 Qy 650 TGGTATTCATCTCCAGAGCGATGAAACCGTTTCAGTTTGTCTCATGGAACCGGTGTAACAA 709
 Db 1341 TGGTATTCATCTCCAGAGCGATGAAACCGTTTCAGTTTGTCTCATGGAACCGGTGTAACAA 1282
 Qy 710 GGGTGAACACTATCCCATATCACCAGCTCAGCTCTTTTCATTTGCCATAGTAATTCGGA 769
 Db 1281 GGGTGAACACTATCCCATATCACCAGCTCAGCTCTTTTCATTTGCCATAGTAATTCGGA 1222
 Qy 770 TGAGCATTTTCATCAGCGGGCAAGATGTAATAAGCGCGGATAAACTTGTGCTTATTT 829
 Db 1221 TGAGCATTTTCATCAGCGGGCAAGATGTAATAAGCGCGGATAAACTTGTGCTTATTT 1162
 Qy 830 TTCTTTACGGTCTTTTAAAGCCCGTAAATATCCAGCTGAACCGGTCTGTTATAGTACAT 889
 Db 1161 TTCTTTACGGTCTTTTAAAGCCCGTAAATATCCAGCTGAACCGGTCTGTTATAGTACAT 1102
 Qy 890 TGAGCAACTGACTGAAATGCTTCAAAATGTTTTCATGATGCCATGGGATATATCAACG 949
 Db 1101 TGAGCAACTGACTGAAATGCTTCAAAATGTTTTCATGATGCCATGGGATATATCAACG 1042

Db 2734 ACCTGCCGATCAAGTCTCATTTTCCGCAAAAGTTGGCCAGGGCTTCCCGGTATCAACA 2675
 QY 1130 GGCACACCGAGTTATTTATTTCTGCAAGTGTATCTTCGTCACAGGTATTTATTCGGTTC 1189
 Db 2674 GGCACACCGAGTTATTTATTTCTGCAAGTGTATCTTCGTCACAGGTATTTATTCGGTTC 2615
 QY 1190 GAAAGGATCGTTTTCGATTTATCTGTAAGTGTATCTTCGTCACAGGTATTTATTCGGTTC 1225
 Db 2614 GAAAGGATCTAGTGAAGATCTCTTTTGTATATCT 2579

RESULT 8

ADD14893/c
 ID ADD14893 standard; DNA; 3817 BP.
 XX
 AC ADD14893;
 XX
 DT 15-JAN-2004 (first entry)
 DE Phase display vector pABMbd-1.
 KW ds; adapter-directed display system; phage display;
 KW outer surface protein; homodimerisation domain;
 KW heterodimerisation domain; antigen-binding unit; cell surface receptor;
 KW receptor ligand; cytosolic protein; secreted protein; HA tag; OmpA;
 KW GABABR; GR2; Lpp.
 XX
 OS Synthetic.
 OS Escherichia coli.
 OS Enterobacteria phage M13.
 OS Homo sapiens.
 XX
 PN US2003104355-A1.
 XX
 PD 05-JUN-2003.
 XX
 PF 02-NOV-2001; 2001US-00033399.
 XX
 PR 02-NOV-2001; 2001US-00033399.
 XX
 PA (WANG/) WANG C.
 PA (ZHON/) ZHONG P.
 PA (WANG/) WANG X.
 XX
 PI Wang C, Zhong P, Wang X;
 XX
 DR WPI; 2003-801222/75.
 XX
 PT Adapter-directed display system having expression vector with sequence
 PT encoding exogenous polypeptide and helper vector, for displaying
 PT exogenous polypeptide e.g., receptor ligand on outer surface of genetic
 PT package.
 XX
 PS Example 6; SEQ ID NO 24; 77pp; English.

XX The invention relates to an adapter-directed display system for
 CC displaying an exogenous polypeptide on the outer surface of a genetic
 CC package, comprising an expression vector with a sequence that encodes
 CC exogenous polypeptide fused to a first adapter sequence, and a helper
 CC vector with outer-surface sequences encoding proteins fused to a second
 CC adapter, the polypeptide is produced in a host cell to cause display of
 CC the polypeptide. Also included are a helper vector for displaying a
 CC polypeptide on the outer surface of a genetic package (comprising, outer-
 CC surface sequences necessary for packaging the genetic package, where at
 CC least one of the surface presenting sequences is fused in-frame to an
 CC adapter, the adapter acting, when the polypeptide is produced in a
 CC suitable host cell, to cause the display of the polypeptide), an
 CC expression vector for producing a polypeptide within or on the outer
 CC surface of a genetic package (comprising a coding sequence encoding the
 CC polypeptide fused in-frame to a first adapter, where the vector is devoid
 CC of outer-surface sequences encoding any functional outer-surface proteins
 CC of the genetic package, and expression of the polypeptide on the outer
 CC surface of the genetic package is mediated through non-covalent pairwise

CC interaction between the first adapter and a second adapter, where the
 CC second adapter is fused to an outer-surface protein), a kit comprising
 CC above vectors in suitable packaging, a host cell comprising the vectors,
 CC a polypeptide displayed on the outer surface of a genetic package using
 CC the adapter-directed display system, a genetic package displaying on its
 CC outer surface the fusion polypeptide, a selectable library comprising
 CC several genetic packages, at least one being the expression vector, a
 CC selectable library comprising several genetic packages (at least one
 CC member of several packages displaying a polypeptide on its outer surface
 CC using the adapter-directed display system) and detecting the presence of
 CC a specific interaction between a test agent and an exogenous polypeptide
 CC that is displayed on a genetic package (involving providing a genetic
 CC package displaying the exogenous polypeptide that is prepared using the
 CC adapter-directed display system, contacting the genetic package with the
 CC test agent and detecting the formation of the stable polypeptide-agent
 CC complex on the genetic package). The outer-surface sequences are chosen
 CC from gene III, gene VI, gene VII, gene VIII and gene IX of a filamentous
 CC phage. The outer-surface sequences encode bacterial outer-surface
 CC proteins chosen from Lpp-OmpA, Trar, Pal, OprI, and AIDA-1. The first
 CC and second adapters are homodimerisation or heterodimerisation sequences.
 CC The adapter-directed display system is useful for displaying a
 CC polypeptide on the outer surface of a genetic package, which involves
 CC causing the adapter-directed display system to be transcribed and
 CC translated in a suitable host cell. The selectable library is useful for
 CC obtaining a polypeptide with desired property, which involves providing
 CC the selectable library and screening the selectable library to obtain at
 CC least one genetic package displaying a polypeptide with the desired
 CC property. The desired property is binding specificity to an agent of
 CC interest. Screening the selectable library further involves isolating the
 CC genetic package that displays a polypeptide having the desired property.
 CC Isolating the genetic package further involves obtaining a nucleotide
 CC sequence from the genetic package that encodes the polypeptide with the
 CC desired property. The polypeptide with the desired property is chosen
 CC from antigen-binding unit, cell surface receptor, receptor ligand,
 CC cytosolic protein, secreted protein, nuclear protein, and their
 CC functional motif. The present sequence represents a vector of the
 CC invention which can express an exogenous protein as a fusion protein with
 CC an Lpp-OmpA-GR2 fusion protein using GABAB receptor GR2 domain as the
 CC adapter sequence.

XX
 SQ Sequence 3817 BP; 948 A; 930 C; 948 G; 991 T; 0 U; 0 Other;

Query Match 90.2%; Score 1130.8; DB 10; Length 3817;
 Best Local Similarity 99.7%; Pred. No. 0;
 Matches 1143; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 50 AAACGGATCTATCGTCAATTTATTTACCTCCACGGGGAGAGCCTGAGCAAACTGCCTCA 109
 Db 2834 AAAAGGATCATATCGTCAATTTATTTACCTCCACGGGGAGAGCCTGAGCAAACTGCCTCA 2775
 QY 110 GGCATTTGAGAAGCACACGGTCACACTGCTTCCGGTAGTCAATAAACCGGTAAACACGCA 169
 Db 2774 GGCATTTGAGAAGCACACGGTCACACTGCTTCCGGTAGTCAATAAACCGGTAAACACGCA 2715
 QY 170 ATAGACATAAGCGGCTATTTAAACGCCCTCGCTGAAACCGACGACCGGGTGAATTTGCT 229
 Db 2714 ATAGACATAAGCGGCTATTTAAACGCCCTCGCTGAAACCGACGACCGGGTGAATTTGCT 2655
 QY 230 TTGGAATTTGCGCATTCATCGGCTTATATATCACTTTATTCAGGCGTAGCAACACGCGTT 289
 Db 2654 TTGGAATTTGCGCATTCATCGGCTTATATCACTTTATTCAGGCGTAGCAACACGCGTT 2595
 QY 290 TAAGGGCAACCAATTAACCTGCTTTAAAAAATTAACGCCCGCCCTGCACCTCATCGAGTAC 349
 Db 2594 TAAGGGCAACCAATTAACCTGCTTTAAAAAATTAACGCCCGCCCTGCACCTCATCGAGTAC 2535
 QY 350 TGTGTAATTCATTAAGCATTTCTGCCGACATGGAAGCCATCAAAAACGGCATGTGAACC 409
 Db 2534 TGTGTAATTCATTAAGCATTTCTGCCGACATGGAAGCCATCAAAAACGGCATGTGAACC 2475
 QY 410 TGAATCGCCAGCGGCATCAGCACCTTGTGCGCTTATATATTTGCCCATGTGAAA 469
 Db 2474 TGAATCGCCAGCGGCATCAGCACCTTGTGCGCTTATATATTTGCCCATGTGAAA 2415

| | | | |
|----|------|---|------|
| Qy | 470 | ACGGGGCGCGAAGAAGTTGTCCATATTGGCCACGTTTAAATCAAAACTGGTGAAACTCAAC | 529 |
| Db | 2414 | AC-GGGCGGAAGAAGTTGTCCATATTGGCCACGTTTAAATCAAAACTGGTGAAACTCAAC | 2356 |
| Qy | 530 | CAGGGATTGGCTGAGACGAAACACATATTTCTCAATAAACCCCTTTAGGGAATAAGGCCAGG | 589 |
| Db | 2355 | CAGGGATTGGCTGAGACGAAACACATATTTCTCAATAAACCCCTTTAGGGAATAAGGCCAGG | 2296 |
| Qy | 590 | TTTTCCCGCTAAACACGCGCACATCTTGCGAATATATGTAGAAACATGCGCGAAATCGTCG | 649 |
| Db | 2295 | TTTTCCCGCTAAACACGCGCACATCTTGCGAATATATGTAGAAACATGCGCGAAATCGTCG | 2236 |
| Qy | 650 | TGGTATTTCACCTCCAGAGCGATGAAAAAGTTTCAGTTTGTCTCATCGAAAAACGGTGTAAACAA | 709 |
| Db | 2235 | TGGTATTTCACCTCCAGAGCGATGAAAAAGTTTCAGTTTGTCTCATCGAAAAACGGTGTAAACAA | 2176 |
| Qy | 710 | GGGTGAACATATCCCATATCACAGCTCACCGTCTTTTCATATGCGCATACGTAATTTCCGGA | 769 |
| Db | 2175 | GGGTGAACATATCCCATATCACAGCTCACCGTCTTTTCATATGCGCATACGTAATTTCCGGA | 2116 |
| Qy | 770 | TGACCATTCATCAGCGCGGCAGNAATGTGAATAAAGCCCGGATAAAACTTGTGCTTATT | 829 |
| Db | 2115 | TGACCATTCATCAGCGCGGCAGNAATGTGAATAAAGCCCGGATAAAACTTGTGCTTATT | 2056 |
| Qy | 830 | TTCTTTTACGGTCTTTAAAAAGGCCGTAAATATCCAGCTGAAACGGTCTGTGTTATAGGTACAT | 889 |
| Db | 2055 | TTCTTTTACGGTCTTTAAAAAGGCCGTAAATATCCAGCTGAAACGGTCTGTGTTATAGGTACAT | 1996 |
| Qy | 890 | TGACCAACTGAATGAAATGCGTCAAAATGTTCTTTTACGATGCCATTTGGGATATATCAACG | 949 |
| Db | 1995 | TGACCAACTGAATGAAATGCGTCAAAATGTTCTTTTACGATGCCATTTGGGATATATCAACG | 1936 |
| Qy | 950 | GTGGTATATCCAGTGATTTTTTTCTCCATTTTAGCTTCCTTAGCTCTGAAAAATCTCGAC | 1009 |
| Db | 1935 | GTGGTATATCCAGTGATTTTTTTCTCCATTTTAGCTTCCTTAGCTCTGAAAAATCTCGAC | 1876 |
| Qy | 1010 | AACCTCAAAAATACGCCCGGTAGTGATCTTATTTTCAATTATGGTGAAAGTTGGAAACCTCTT | 1069 |
| Db | 1875 | AACCTCAAAAATACGCCCGGTAGTGATCTTATTTTCAATTATGGTGAAAGTTGGAAACCTCTT | 1816 |
| Qy | 1070 | ACGTGCCGATCAACGCTCTCATTTTTTCGCCAAAAAGTTGGCCACAGGCTTCCCGGTATCAACA | 1129 |
| Db | 1815 | ACGTGCCGATCAACGCTCTCATTTTTTCGCCAAAAAGTTGGCCACAGGCTTCCCGGTATCAACA | 1756 |
| Qy | 1130 | GGGACACCGAGTTTATTTATTCCTCGGAAGTGATCTTTCGGTCAAGGATTTTATTCGGTC | 1189 |
| Db | 1755 | GGGACACCGAGTTTATTTATTCCTCGGAAGTGATCTTTCGGTCAAGGATTTTATTCGGTC | 1696 |
| Qy | 1190 | GAAGAAG 1195 | |
| Db | 1695 | GAAGAAG 1690 | |

[illegible]

QY 650 TGGTATTTCACTCCAGAGCGATGAACAGTTTCAGTTTGCTCATGGAACGGTGTAAACA 709
 Db 4262 TGGTATTTCACTCCAGAGCGATGAACAGTTTCAGTTTGCTCATGGAACGGTGTAAACA 4203
 QY 710 GGGTGAACACTATCCCATATACACAGCTCACCGTCTTTTCATTTGCCATAGCTAATTCGGA 769
 Db 4202 GGGTGAACACTATCCCATATACACAGCTCACCGTCTTTTCATTTGCCATAGCTAATTCGGA 4143
 QY 770 TGAGCATTCATCAGCGGGGCAAGATGTGAATAAAGCGCGGATAAACCTTTGCTTATTT 829
 Db 4142 TGAGCATTCATCAGCGGGGCAAGATGTGAATAAAGCGCGGATAAACCTTTGCTTATTT 4083
 QY 830 TTCTTTACGGTCTTTTAAAGCGCGTAAATATCCAGCTGAACGGTCTGGTTATAGGTACAT 889
 Db 4082 TTCTTTACGGTCTTTTAAAGCGCGTAAATATCCAGCTGAACGGTCTGGTTATAGGTACAT 4023
 QY 890 TGAGCAACTGACTGAAATGCCTCAAAATGCTTTTACGATGCATTTGGGATATATCAACG 949
 Db 4022 TGAGCAACTGACTGAAATGCCTCAAAATGCTTTTACGATGCATTTGGGATATATCAACG 3963
 QY 950 GTGGTATATCCAGTGATTTTTTCTCCATTTTTCAGTCTTCTAGCTTCTGAAATCTCGAC 1009
 Db 3962 GTGGTATATCCAGTGATTTTTTCTCCATTTTTCAGTCTTCTAGCTTCTGAAATCTCGAT 3903
 QY 1010 AACTCAAAAATACGCGCGTAGTGATCTTATTTCAATATGTTGAAAGTTGGAACCTCTT 1069
 Db 3902 AACTCAAAAATACGCGCGTAGTGATCTTATTTCAATATGTTGAAAGTTGGAACCTCTT 3843
 QY 1070 AGTGGCCGATCAACGCTCTCATTTTCGCCAAAAGTTGGCCAGGGCTTCCCGGTATCAACA 1129
 Db 3842 AGTGGCCGATCAACGCTCTCATTTTCGCCAAAAGTTGGCCAGGGCTTCCCGGTATCAACA 3783
 QY 1130 GGGACACCAAGTATTTATTTCTGGAAGTGATCTTCGGTCAACAGGTATTTATTCGGTTC 1189
 Db 3782 GGGACACCAAGTATTTATTTCTGGAAGTGATCTTCGGTCAACAGGTATTTATTCGGTTC 3723
 QY 1190 GAAAGGATCCG 1201
 Db 3722 CAAAGTGGCTCG 3711

RESULT 10
 ADZ03870/c
 ID ADZ03870 standard; DNA; 6608 BP.
 AC ADZ03870;
 XX
 DT 16-JUN-2005 (first entry)
 XX
 DE Plasmid P15a DNA sequence #2.
 XX
 KW Pharmaceutical; P15a; ds.
 XX
 OS Plasmid P15a.
 XX
 PN WO2005030972-A1.
 XX
 PD 07-APR-2005.
 XX
 PF 29-SEP-2004; 2004WO-JP014213.
 XX
 PR 29-SEP-2003; 2003JP-00337663.
 XX
 PA (JAPC) NIPPON SHOKUBAI CO LTD.
 XX
 PI Toraya T, Tobimatsu T, Yamanishi M, Mori K, Kajiuura H, Yamada S;
 PI Yuzuki M, Azuma M, Hara T, Yasuda S;
 XX
 DR WPI; 2005-315044/32.
 XX
 PT Manufacture of 3-hydroxy propionaldehyde useful as raw material for
 PT manufacturing 1,3-propanediol, by dehydrating glycerol with diol
 PT dehydratase and/or glycerol dehydratase and microbial cells.

XX Example 2; SEQ ID NO 4; 40pp; Japanese.
 PS
 XX
 CC The invention relates to a method for manufacturing 3-
 CC hydroxypropionaldehyde involving dehydrating glycerol with diol
 CC dehydratase and/or glycerol dehydratase and microbial cells comprising
 CC diol dehydratase reactivation factor and/or glycerol dehydratase
 CC reactivation factor. The invention also relates to a method for
 CC manufacturing 1,3-propanediol, a method for manufacturing 3-
 CC hydroxypropionic acid and methods for manufacturing acrolein, acrylic
 CC acid and acrylic ester. The dehydration reaction of glycerol is performed
 CC under aerobic conditions using microbial cells. This sequence represents
 CC plasmid P15a DNA used in the method of the invention.
 XX
 SQ Sequence 6608 BP; 1495 A; 1749 C; 1918 G; 1446 T; 0 U; 0 Other;
 Query Match 89.0%; Score 1116; DB 14; Length 6608;
 Best Local Similarity 98.6%; Pred. No. 0;
 Matches 1136; Conservative 0; Mismatches 15; Indels 1; Gaps 1;
 QY 50 AAAACGGATCTATCGTCAATTTATTTACCTCCACGGGAGAGCCTGAGCAAACTGSCCTCA 109
 Db 4862 ATAAATGATCATATCGTCAATTTATTTACCTCCACGGGAGAGCCTGAGCAAACTGSCCTCA 4803
 QY 110 GGCATTTGAGAAGCACACGGTCACTGCTTCCGCTAGTCAATAAACCGGTAAACAGCA 169
 Db 4802 GGCATTTGAGAAGCACACGGTCACTGCTTCCGCTAGTCAATAAACCGGTAAACAGCA 4743
 QY 170 ATAGACATAACGGCTATTTAACGACCTGCTGACCGACGACCGCGGTGCAATTTGCT 229
 Db 4742 ATAGACATAACGGCTATTTAACGACCTGCTGACCGACGACCGCGGTGCAATTTGCT 4683
 QY 230 TTCGAATTTTCGCCATTCATCCGGCTTATTTACCTTATTTACGGGTAGCAACCGCGCTT 289
 Db 4682 TTCGAATTTTCGCCATTCATCCGGCTTATTTACCTTATTTACGGGTAGCAACCGCGCTT 4624
 QY 290 TAAGGGACCAATTAACGCTTTAAAAAATTAACGCGCCGCTGCACTCATCGCAGTAC 349
 Db 4623 TAAGGGACCAATTAACGCTTTAAAAAATTAACGCGCCGCTGCACTCATCGCAGTAC 4564
 QY 350 TGTGTATTTCAATTAAGCATTTCTCCGACATGGAAGCCATCACAGACGGCATGTAAC 409
 Db 4563 TGTGTATTTCAATTAAGCATTTCTCCGACATGGAAGCCATCACAGACGGCATGTAAC 4504
 QY 410 TGAATTCGCCAGCGCATCAGCACCTTTGCGCTTTCGCTATTAATTTTCCCATGCTGAAA 469
 Db 4503 TGAATTCGCCAGCGCATCAGCACCTTTGCGCTTTCGCTATTAATTTTCCCATGCTGAAA 4444
 QY 470 ACGGGGCGAAGAGTTGTTCATATTTGCCACGTTTAAATCAAACTCGTGAACCTCAC 529
 Db 4443 ACGGGGCGAAGAGTTGTTCATATTTGCCACGTTTAAATCAAACTCGTGAACCTCAC 4384
 QY 530 CAGGGATTTGCTGAGACGAAAAACATATTTCTCAATAAACCTTTAGGAAATAGCCAGG 589
 Db 4383 CAGGGATTTGCTGAGACGAAAAACATATTTCTCAATAAACCTTTAGGAAATAGCCAGG 4324
 QY 590 TTTTCACCGTAACACGCGCATCTTTCGCAATATATGTAGAAAACCTGCGGAAATCGTCG 649
 Db 4323 TTTTCACCGTAACACGCGCATCTTTCGCAATATATGTAGAAAACCTGCGGAAATCGTCG 4264
 QY 650 TGGTATTTCACTCCAGAGCGATGAAAAAGTTTCAGTTTGTCTCATGGAACCGGTGTAAACA 709
 Db 4263 TGGTATTTCACTCCAGAGCGATGAAAAAGTTTCAGTTTGTCTCATGGAACCGGTGTAAACA 4204
 QY 710 GGGTGAACACTATCCCATATACACAGCTCACCGTCTTTTCATTTGCCATAGCTAATTCGGA 769
 Db 4203 GGGTGAACACTATCCCATATACACAGCTCACCGTCTTTTCATTTGCCATAGCTAATTCGGA 4144
 QY 770 TGAGCATTTTCATCAGCGGGCAAGATGTGAATAAAGCGCGGATAAACCTTTGCTTATTT 829
 Db 4143 TGAGCATTTTCATCAGCGGGCAAGATGTGAATAAAGCGCGGATAAACCTTTGCTTATTT 4084
 QY 830 TTCTTTACGGTCTTTTAAAGCGCGTAAATATCCAGCTGAACGGTCTGGTTATAGGTACAT 889

Db 4083 TTCTTTACGGTCTTTAAAAAGGCGCTAATATCCAGCTGAACGGTCTGGTATAGTACAT 4024
QY 890 TGAGCAACTGACTGAAATGCCCTCAAAATGTTCTTTACGATGCCATTTGGATATATCAACG 949
Db 4023 TGAGCAACTGACTGAAATGCCCTCAAAATGTTCTTTACGATGCCATTTGGATATATCAACG 3964
QY 950 GTGGTATATCCAGTCAATTTTCTCCATTTTAGCTTCTTAGCTCCCTGAAATCTCGAC 1009
Db 3963 GTGGTATATCCAGTCAATTTTCTCCATTTTAGCTTCTTAGCTCCCTGAAATCTCGAT 3904
QY 1010 AACTCAAAAAATPACCCCGGTAGTGATCTTATTTCAATTATGTTGAAAGTTGAACTCTTT 1069
Db 3903 AACTCAAAAAATPACCCCGGTAGTGATCTTATTTCAATTATGTTGAAAGTTGAACTCTTT 3844
QY 1070 AGTGCCGATCAACGTCTCAATTTTCGCCAAAGTTGGCCGAGGCTTCCCGGTATCAACA 1129
Db 3843 AGTGCCGATCAACGTCTCAATTTTCGCCAAAGTTGGCCGAGGCTTCCCGGTATCAACA 3784
QY 1130 GGGACACAGGATTTATTTATCTCGAAGTGATCTTCGGTCACAGGTATTTATTCGGTC 1189
Db 3783 GGGACACAGGATTTATTTATCTCGAAGTGATCTTCGGTCACAGGTATTTATTCGGTC 3724
QY 1190 GAAAGAGGATCCG 1201
Db 3723 CAAAGTGGCTCG 3712

RESULT 11
ADZ66224/c
ID ADZ66224 standard; cDNA; 9808 BP.
XX AC ADZ66224;
XX 14-JUL-2005 (first entry)
XX Construct pAM401-MCS.
XX ss; Cytostatic; Antiasthmatic; Respiratory-Gen; Antipsoriatic;
KW Dermatological; Antiangiogenic; Ophthalmological; Antidiabetic;
KW Antirheumatic; Antiarthritic; Antiinflammatory; Immunosuppressive;
KW Uropathic; Gynecological; Antiatheosclerotic; Cardiant; Vasotropic;
KW Vaccine; EphA2; receptor tyrosine kinase; Ephrin; hyperproliferative;
KW Listeria; angiogenesis; cancer; skin; lung; colon; breast; prostate;
KW bladder; pancreas; renal cell carcinoma; melanoma; leukemia; lymphoma;
KW asthma; chronic pulmonary obstructive disease; lung fibrosis;
KW bronchial hyper responsiveness; psoriasis; seborrheic dermatitis;
KW macular degeneration; diabetic retinopathy; retinopathy of prematurity;
KW vascular restenosis; infantile hemangioma; verruca vulgaris; psoriasis;
KW Kaposi's sarcoma; neurofibromatosis;
KW recessive dystrophic epidermolysis bullosa; rheumatoid arthritis;
KW ankylosing spondylitis; systemic lupus; psoriatic arthropathy;
KW Reiter's syndrome; Sjogrens' syndrome; endometriosis; preeclampsia;
KW atherosclerosis; coronary artery disease; gene.
XX Synthetic.
XX OS
XX WO2005037233-A2.
PN
PD 28-APR-2005.
XX
XX 15-OCT-2004; 2004WO-US034694.
XX
PR 15-OCT-2003; 2003US-0511719P.
PR 15-OCT-2003; 2003US-0511919P.
PR 24-DEC-2003; 2003US-0532666P.
PR 26-MAR-2004; 2004US-0556631P.
PR 01-OCT-2004; 2004US-0615470P.
PR 07-OCT-2004; 2004US-0617544P.
XX
XX (MEDI-) MEDIMUNE INC.
PA (CERU-) CERUS CORP.
XX

PI Kinch MS, Kiener PA, Bruckheimer E, Dubensky TW, Cook DN;
XX WPI; 2005-322763/33.
XX
XX Eliciting an immune response against EphA2-expressing cells in subjects,
PT by administering composition comprising Listeria bacterium that expresses
PT EphA2 antigenic peptide, to subject.
XX
XX Disclosure; SEQ ID NO 43; 219pp; English.
XX
XX This sequence represents the expression cassette construct pAM401-MCS.
CC This sequence was used in the scope of the invention for the expression
CC of EphA2. EphA2 is a 130 kD receptor tyrosine kinase that is expressed at
CC low levels in adult epithelia and is enriched within sites of cell-cell
CC adhesion. It binds ligands (Ephrins A1-A5) that are anchored to the cell
CC membrane. Ligand binding causes autophosphorylation, however EphA2
CC retains enzymatic activity in the absence of ligand binding or
CC phosphorylation content. EphA2 is upregulated on a large number of
CC hyperproliferation cells, including aggressive carcinoma cells. The
CC method of the invention for eliciting an immune response against an EphA2
CC -expressing cell in a subject, involves administering to a subject a
CC composition comprising a Listeria bacterium that expresses an EphA2
CC antigenic peptide, to elicit an immune response against an EphA2-
CC expressing cell. The invention contains further methods for treating a
CC human subject having a hyperproliferative disorder of EphA2-expressing
CC cells, involving administering to the subject a composition comprising
CC EphA2 antigen peptide-expressing Listeria bacterium, to treat a
CC hyperproliferative disorder of EphA2-expressing cells; and treating a
CC human subject having a disease involving aberrant angiogenesis involving
CC administering to the subject a composition comprising EphA2 antigenic
CC peptide-expressing Listeria bacterium, to treat disease involving
CC aberrant angiogenesis. The method of the invention is useful for
CC eliciting an immune response against an EphA2-expressing cell in a
CC subject. The subject has cancer. The cancer is of an epithelial cell
CC origin or T cell origin. The cancer is a cancer of skin, lung, colon,
CC breast, prostate, bladder or pancreas or is a renal cell carcinoma or
CC melanoma. The cancer is a leukemia or a lymphoma. The subject has a non-
CC neoplastic hyperproliferative disorder. The hyperproliferative disorder
CC is an epithelial cell disorder. The hyperproliferative disorder is
CC asthma, chronic pulmonary obstructive disease, lung fibrosis, bronchial
CC hyper responsiveness, psoriasis, and seborrheic dermatitis. The second
CC method is useful for treating a human subject having a hyperproliferative
CC disorder of EphA2-expressing cells. The third method is useful for
CC treating a subject having a disease involving aberrant angiogenesis. The
CC disease is macular degeneration, diabetic retinopathy, retinopathy of
CC prematurity, vascular restenosis, infantile hemangioma, verruca vulgaris,
CC psoriasis, Kaposi's sarcoma, neurofibromatosis, recessive dystrophic
CC epidermolysis bullosa, rheumatoid arthritis, ankylosing spondylitis,
CC systemic lupus, psoriatic arthropathy, Reiter's syndrome, and Sjogrens'
CC syndrome, endometriosis, preeclampsia, atherosclerosis or coronary artery
CC disease.
XX
SQ Sequence 9808 BP; 2836 A; 1923 C; 1946 G; 3103 T; 0 U; 0 Other;
Query Match 89.0%; Score 1116; DB 14; Length 9808;
Best Local Similarity 98.6%; Pred. No. 0;
Matches 1136; Conservative 0; Mismatches 15; Indels 1; Gaps 1;
QY 50 AAAACGGATCCTATCGTCAATTTATCTCACGGGAGAGCCTGAGCAAACTGGCCTCA 109
Db 9192 ATAAATGATCATATCGTCAATTTATCTCACGGGAGAGCCTGAGCAAACTGGCCTCA 9133
QY 110 GGCATTTTGAGAGCACACGGTCACTGCTTCGGTAGTCAATAACCGTAACACGCA 169
Db 9132 GGCATTTTGAGAGCACACGGTCACTGCTTCGGTAGTCAATAACCGTAACACGCA 9073
QY 170 ATAGACATAAGCGGCTATTTAAACGACCTGCTCCCTGAAACCGACCGGTCGAATTCCT 229
Db 9072 ATAGACATAAGCGGCTATTTAAACGACCTGCTCCCTGAAACCGACCGGTCGAATTCCT 9013
QY 230 TTCGAATTTTCGCCATTTCATCCGCTTATATCACTTATTCAGGGGTAGCAACCGCGCTT 289
Db 9012 TTCGAATTTTCGCCATTTCATCCGCTTATATCACTTATTCAGGGGTAGC-ACCGCGCTT 8954

| | | | | | | | | | | | | | |
|----|------|----------|-----------|--------------|---------|---------|-----------|----------|----------|--------------|---------------|----------------|-------------|
| Qy | 290 | TAAGGGCA | CCAAATAA | CTGCCTTAAAAA | AAATTA | CGCCCCG | CGCTTGCCA | CTCATCG | CAGTAC | 349 | | | |
| Db | 8953 | TAAGGGCA | CCAAATAA | CTGCCTTAAAAA | AAATTA | CGCCCCG | CGCTTGCCA | CTCATCG | CAGTAC | 8894 | | | |
| Qy | 350 | TGTTGTA | TTTCAATTA | AGCATTCT | GC | CGACAT | TGG | AGCCAT | CACAAAC | CGGCATGATGA | CC 409 | | |
| Db | 8893 | TGTTGTA | TTTCAATTA | AGCATTCT | GC | CGACAT | TGG | AGCCAT | CACAAAC | CGGCATGATGA | CC 8834 | | |
| Qy | 410 | TGAATCG | CCACGCG | CATCAG | CACCTT | GC | CGCTTG | CGCTATA | ATAATTTT | TGCCCATGGTGA | AA 469 | | |
| Db | 8833 | TGAATCG | CCACGCG | CATCAG | CACCTT | GC | CGCTTG | CGCTATA | ATAATTTT | TGCCCATGGTGA | AA 8774 | | |
| Qy | 470 | ACGGGGC | GGAAGATT | TG | TCATAT | TGG | CCA | CGTTTAA | ATCAAA | ATCTGGTGA | AACTCACC 529 | | |
| Db | 8773 | ACGGGGC | GGAAGATT | TG | TCATAT | TGG | CCA | CGTTTAA | ATCAAA | ATCTGGTGA | AACTCACC 8714 | | |
| Qy | 530 | CAGGGAT | TGGCTG | AGACG | AAACAT | ATTTCT | CAATA | AAACCTTT | TAG | GGAAATAGG | CCAGG 589 | | |
| Db | 8713 | CAGGGAT | TGGCTG | AGACG | AAACAT | ATTTCT | CAATA | AAACCTTT | TAG | GGAAATAGG | CCAGG 8654 | | |
| Qy | 590 | TTTTTCA | CCGTAA | CACG | CCACAT | CTTG | CGAATAT | ATGTGA | AACTCC | CGGAATCG | TCG 649 | | |
| Db | 8653 | TTTTTCA | CCGTAA | CACG | CCACAT | CTTG | CGAATAT | ATGTGA | AACTCC | CGGAATCG | TCG 8594 | | |
| Qy | 650 | TGTTATT | CACTCC | AGAG | CGATG | AAAA | CGTTT | CAGTTT | GC | CATGG | AAAAACGGTGA | CAAA 709 | |
| Db | 8593 | TGTTATT | CACTCC | AGAG | CGATG | AAAA | CGTTT | CAGTTT | GC | CATGG | AAAAACGGTGA | CAAA 8534 | |
| Qy | 710 | GGGTGA | ACATAT | CCCAT | ATAC | CAG | CTCAC | CGCTTTT | TTCAT | TG | CCATACGTA | ATCTCCGGA 769 | |
| Db | 8533 | GGGTGA | ACATAT | CCCAT | ATAC | CAG | CTCAC | CGCTTTT | TTCAT | TG | CCATACGTA | ATCTCCGGA 8474 | |
| Qy | 770 | TGACAT | TTCAT | CAGG | CGGCA | GNAT | TGTA | ATAAG | CCG | GATAA | ACTTTG | TGCTTATTT 829 | |
| Db | 8473 | TGACAT | TTCAT | CAGG | CGGCA | GNAT | TGTA | ATAAG | CCG | GATAA | ACTTTG | TGCTTATTT 8414 | |
| Qy | 830 | TTCTTTA | CGCTCT | TTTAAAA | AGGCG | TAATAT | CCAG | CTG | AAACGG | CTCTG | TTTATAGG | TACAT 889 | |
| Db | 8413 | TTCTTTA | CGCTCT | TTTAAAA | AGGCG | TAATAT | CCAG | CTG | AAACGG | CTCTG | TTTATAGG | TACAT 8354 | |
| Qy | 890 | TGACAA | CTGACT | GAAT | TGCCT | CAAA | ATGTTCTTT | TAC | GATGCC | ATTTGG | GATATATCA | ACG 949 | |
| Db | 8353 | TGACAA | CTGACT | GAAT | TGCCT | CAAA | ATGTTCTTT | TAC | GATGCC | ATTTGG | GATATATCA | ACG 8294 | |
| Qy | 950 | GTGGTA | TATCC | AGTGAT | TTTTTTT | CT | CAATTTT | TAG | CTTCTT | TAGCTC | TCTG | AAAAATCTCG | AC 1009 |
| Db | 8293 | GTGGTA | TATCC | AGTGAT | TTTTTTT | CT | CAATTTT | TAG | CTTCTT | TAGCTC | TCTG | AAAAATCTCG | AC 8234 |
| Qy | 1010 | AACTCA | AAAAAT | AGCC | CGGTAG | TGAT | CTTATTT | CA | TATGGT | CAAA | AGTTTGG | AACTCTTT 1069 | |
| Db | 8233 | AACTCA | AAAAAT | AGCC | CGGTAG | TGAT | CTTATTT | CA | TATGGT | CAAA | AGTTTGG | AACTCTTT 8174 | |
| Qy | 1070 | ACGTG | CCGATCA | AGCTCT | CAATTTT | TG | CCAAAA | AGTTTGG | CC | AGG | CTTCC | CGGTATCA | ACA 1129 |
| Db | 8173 | ACGTG | CCGATCA | AGCTCT | CAATTTT | TG | CCAAAA | AGTTTGG | CC | AGG | CTTCC | CGGTATCA | ACA 8114 |
| Qy | 1130 | GGGAC | ACCAG | ATTTATTT | ATTTCT | TG | CGAAGT | GTATCTT | CCG | TC | CAGG | ATTTATTT | TCGGTC 1189 |
| Db | 8113 | GGGAC | ACCAG | ATTTATTT | ATTTCT | TG | CGAAGT | GTATCTT | CCG | TC | CAGG | ATTTATTT | TCGGTC 8054 |
| Qy | 1190 | GA | AAAG | ATCG | 1201 | | | | | | | | |
| Db | 8053 | CA | AA | ATCG | TCG | 8042 | | | | | | | |

RESULT 12
AEB80052/c
ID AEB80052 standard; DNA; 9808 BP.
XX
AC AEB80052;
XX
DT 06-OCT-2005 (first entry)

Vector pAM401/pPL2 containing hly promoter/p60 sig pep. cassette.

Vector; ds; listeriolysin O; Cytostatic; Antimicrobial; Immunosuppressive; Immunostimulant; vaccine; fusion protein; protein production; protein secretion; immune response; cancer; infection; autoimmune disease.

Listeria monocytogenes.

Synthetic.

Unidentified.

W02005071088-A2.

04-AUG-2005.

23-DEC-2004; 2004WO-US044080.

24-DEC-2003; 2003US-0532598P.

02-FEB-2004; 2004US-0541515P.

06-FEB-2004; 2004US-00773618.

06-FEB-2004; 2004US-00773792.

26-MAR-2004; 2004US-0556744P.

30-JUN-2004; 2004US-00883599.

23-JUL-2004; 2004WO-US023881.

05-AUG-2004; 2004US-0599377P.

01-OCT-2004; 2004US-0615287P.

06-OCT-2004; 2004US-0616750P.

(CERU-) CERUS CORP.

Dubensky TW, Portnoy DA, Luckett WS, Cook DN; WPI; 2005-534010/54.

New recombinant nucleic acid molecule (comprising a first polynucleotide encoding a signal peptide native to a bacterium; and a second polynucleotide encoding a polypeptide) useful e.g. to induce immune response in host to antigen.

Example 13; SEQ ID NO 33; 27lpp; English.

The invention relates to a recombinant nucleic acid molecule (I) (comprising a first polynucleotide (A) encoding a signal peptide native to a bacterium, where (A) is codon-optimized for expression in the bacterium, and a second polynucleotide (B) encoding a polypeptide, where (B) is in the same translational reading frame as (A), where (A) encodes a fusion protein comprising the signal peptide and the polypeptide). Also included are the expression cassette (I) further comprising a promoter operably linked to (A) and (B), a recombinant *Listeria* bacterium (II) comprising (I), an immunogenic composition or vaccine comprising (II), a method of preventing or treating a condition in a host (comprising administering the recombinant *Listeria* bacterium to the host), a recombinant *Listeria* bacterium comprising a polycistronic expression cassette. The polypeptide encoded by (B) comprises an antigen (a tumor-associated antigen, a polypeptide derived from a tumor-associated antigen, an infectious disease antigen, or a polypeptide derived from an infectious disease antigen). The polypeptide encoded by (B) is heterologous to the signal peptide, foreign to the *Listeria* bacterium, or both. The signal peptide is an LLO (*Listeriolysin O*, encoded by the hly gene) signal peptide from *Listeria* monocytogenes or is a p60 signal peptide from *Listeria* monocytogenes. The heterologous protein is preferably mesothelin, an antigenic fragment or its variants, NY-ESO-1, or an antigenic fragment or its variant). The polypeptide encoded by (B) comprises human mesothelin deleted of its signal peptide and GPI linker domain. The recombinant bacterium is useful in the manufacture of a medicament for inducing an immune response in a host to an antigen, and for preventing or treating a condition in a host (where the condition is cancer, infections and autoimmunity). The recombinant nucleic acid molecule, expression cassettes and vectors are useful for expressing and/or secreting polypeptides including heterologous polypeptides such as antigens in bacteria. The present sequence represents the Vector pAM401 which contains the hly promoter/p60 signal peptide cassette.

XX SQ Sequence 9808 BP; 2836 A; 1923 C; 1946 G; 3103 T; 0 U; 0 Other;
Query Match 89.0%; Score 1116; DB 14; Length 9808;
Best Local Similarity 98.6%; Pred. No. 0;
Matches 1136; Conservative 0; Mismatches 15; Indels 1; Gaps 1;
QY 50 AAAACGGATCCTATCGTCAATTTATTACCTCCACGGGAGAGCCTGAGCAAACTGCGCTCA 109
DB 9192 ATAAATGATCATATCGTCAATTTATTACCTCCACGGGAGAGCCTGAGCAAACTGCGCTCA 9133
QY 110 GGCATTTGAGAGACACACGGTCACACTGCTTCCGGTAGTCAATAAACCGGTAAACCCAGCA 169
DB 9132 GGCATTTGAGAGACACACGGTCACACTGCTTCCGGTAGTCAATAAACCGGTAAACCCAGCA 9073
QY 170 ATAGACATAAGCGGTATTTAAAGACCTTCCCTGAACCGAGACCGGGTGAATTTGCT 229
DB 9072 ATAGACATAAGCGGTATTTAAAGACCTTCCCTGAACCGAGACCGGGTGAATTTGCT 9013
QY 230 TTCGAATTTGCGCAATTCATCGCTTATTATCACTTATTCAGGGCTAGCAACCGAGCGTT 289
DB 9012 TTCGAATTTGCGCAATTCATCGCTTATTATCACTTATTCAGGGCTAGC-ACCAGGCGTT 8954
QY 290 TAAGGGCCACCAATACTGCTTTAAAAAATAATACGCGCGCCCTGCGCACTCATCGAGTAC 349
DB 8953 TAAGGGCCACCAATACTGCTTTAAAAAATAATACGCGCGCCCTGCGCACTCATCGAGTAC 8894
QY 350 TGTGTAATTTCAATTAAGCATTTCTGCGACATGGAAGCCATCACAAACCGCATGATGAACC 409
DB 8893 TGTGTAATTTCAATTAAGCATTTCTGCGACATGGAAGCCATCACAGCGCATGATGAACC 8834
QY 410 TCAATCGCCAGGGCATCAGACCTTGTGCGCTTGGGTATATATTTGCCCATGTTGAAA 469
DB 8833 TGAATCGCCAGGGCATCAGACCTTGTGCGCTTGGGTATATATTTGCCCATGTTGAAA 8774
QY 470 ACGGGGGCGAAGAGTTGTCCATATTGGCCAGCTTTAAATCAAAACTGGTGAAACTCAC 529
DB 8773 ACGGGGGCGAAGAGTTGTCCATATTGGCCAGCTTTAAATCAAAACTGGTGAAACTCAC 8714
QY 530 CAGGATTTGGCTGAGACGAAACATATTTCTCAATAAACCCCTTTAGGGAATATAGGCCAGG 589
DB 8713 CAGGATTTGGCTGAGACGAAACATATTTCTCAATAAACCCCTTTAGGGAATATAGGCCAGG 8654
QY 590 TTTTCACCGTAAACGCGACATCTTGGCAATATATGTAGAACTGCGGGAATTCGTCG 649
DB 8653 TTTTCACCGTAAACGCGACATCTTGGCAATATATGTAGAACTGCGGGAATTCGTCG 8594
QY 650 TGGTATTTCACTCCAGAGCGATGAAAAAGTTTCAGTTTGTCTATGAAAAACGGTGTAAACA 709
DB 8593 TGGTATTTCACTCCAGAGCGATGAAAAAGTTTCAGTTTGTCTATGAAAAACGGTGTAAACA 8534
QY 710 GGGTGAACACTATCCCATATACAGCTACCGTCTTTTCAATGCGCATACGTAATTCGGA 769
DB 8533 GGGTGAACACTATCCCATATACAGCTACCGTCTTTTCAATGCGCATACGTAATTCGGA 8474
QY 770 TGAGCATTTATCAGCGGGCAAGATGGAATAAAGCGCGATATAAACTTGCTTATTT 829
DB 8473 TGAGCATTTATCAGCGGGCAAGATGGAATAAAGCGCGATATAAACTTGCTTATTT 8414
QY 830 TTTCTTTAGGCTTTTAAAAAGCGCGTAAATATCCAGCTGAACGGTCTGGTTATAGTACAT 889
DB 8413 TTTCTTTAGGCTTTTAAAAAGCGCGTAAATATCCAGCTGAACGGTCTGGTTATAGTACAT 8354
QY 890 TGAGCAACTGACTGAAATGCTTCAAAATGTTCTTTTACGATGCAATGGGATATATCAACG 949
DB 8353 TGAGCAACTGACTGAAATGCTTCAAAATGTTCTTTTACGATGCAATGGGATATATCAACG 8294
QY 950 GTGGTATATCAGTGAATTTTTTCTCCATTTTACGTTCTTACGTTCTTGAATAATTCGAC 1009
DB 8293 GTGGTATATCAGTGAATTTTTTCTCCATTTTACGTTCTTACGTTCTTGAATAATTCGAT 8234
QY 1010 AACTCAAAAAATACGCCGGTAGTGCATCTTATTTCAATATGTTGGAAGTTGGAACCTCTT 1069

DB 8233 AACTCAAAAAATACGCCGGTAGTGCATCTTATTTTATTATGTTGAAAGTTGGAACCTCTT 8174
QY 1070 ACGTGCCGATCAACGTCCTCATTTTCGCAAAAGTTGCCAGGGCTTCCCGGTATCAACA 1129
DB 8173 ACGTGCCGATCAACGTCCTCATTTTCGCAAAAGTTGCCAGGGCTTCCCGGTATCAACA 8114
QY 1130 GGGACACACAGGATTTATTTATTTCTGCAAGTGATCTTCCGTCACAGGTATTTATTCGTC 1189
DB 8113 GGGACACACAGGATTTATTTATTTCTGCGAAGTGATCTTCCGTCACAGGTATTTATTCGCG 8054
QY 1190 GAAAAAGATCG 1201
DB 8053 CAAAAGTGGTCG 8042
RESULT 13
ID AEB72933 standard; DNA; 9808 BP.
XX AC AEB72933;
XX DT 06-OCT-2005 (first entry)
XX DE DNA sequence for plasmid pAM401-MCS.
XX KW immune response; receptor tyrosine kinase; EphA2; antigen; expression;
KW antibody production; adoptive immunotherapy; vaccine; hyperproliferation;
KW cancer; angiogenesis disorder; immunostimulant; cytostatic;
KW antiangiogenic; antiasthmatic; respiratory-gen; antipsoriatic;
KW antisporadic; dermatological; ophthalmological; antidiabetic;
KW antirheumatic; antiarthritic; antiinflammatory; gynecological;
KW angiogenesis-inhibitor; vasotropic; neuroprotective; immunosuppressive;
KW antiarteriosclerotic; cardiant; pAM401-MCS; ds.
XX OS Synthetic.
XX OS Unidentified.
XX PN WO2005067460-A2.
XX PD 28-JUL-2005.
XX PF 15-OCT-2004; 2004WO-US034693.
XX PR 24-DEC-2003; 2003US-0532696P.
PR 18-AUG-2004; 2004US-0602588P.
PR 01-OCT-2004; 2004US-0615548P.
PR 07-OCT-2004; 2004US-0617564P.
XX (MEDI-) MEDIMUNE INC.
XX PA Kinch MS;
XX PI WPI; 2005-522701/53.
XX PT Eliciting an immune response comprises administering to an individual a
PT composition comprising an EphA2 antigenic peptide, an EphA2 antigenic
PT peptide expression vehicle, or antigen presenting cells.
XX PS Disclosure; SEQ ID NO 43; 241pp; English.
XX CC The invention relates to a method of eliciting an immune response against
CC an EphA2-expressing cell. The method comprises administering to an
CC individual a composition comprising an EphA2 antigenic peptide, an EphA2
CC antigenic peptide expression vehicle, antigen presenting cells sensitized
CC with an EphA2 antigenic peptide, or an anti-idiotypic antibody or its
CC antigen-binding fragment, which immunospecifically binds to an idotype
CC of an anti-EphA2 antibody, in an amount to elicit an immune response
CC against an EphA2-expressing cell. Also described are: (1) a method of
CC treating a human individual having a hyperproliferative disorder of an
CC EphA2-expressing cell, or a disease involving aberrant angiogenesis,
CC comprising performing the method of the invention; and (2) a method
CC of producing antibodies that immunospecifically bind to EphA2, and
CC administering them to an individual as an EphA2 vaccine. In the method of

the invention, the Epha2 antigenic peptide is NOT any of the sequences given as SEQ ID Nos 3-18 in the specification. The composition further comprises an adjuvant, a heat shock protein bound to the Epha2 antigenic peptide, and expression vehicles expressing the Epha2 antigenic peptides. The expression vehicle is an infectious agent comprising a nucleic acid, the nucleic acid comprising a nucleotide sequence encoding the Epha2 antigenic peptide operably linked to a promoter, where the sequence encoding the Epha2 antigenic peptide is codon-optimized for expression in the infectious agent. The method of the invention is useful for treating a human individual having a hyperproliferative disorder of Epha2-expressing cells by administering to the individual a composition comprising antibodies produced by administering an Epha2 vaccine to a host in an amount to treat a hyperproliferative disorder of Epha2-expressing cells. Alternatively, the individual can be administered with a composition comprising an Epha2 expression vehicle, e.g. bacterium or virus, in an amount to treat a hyperproliferative disorder of Epha2-expressing cells, and administering to the individual an antibiotic or antiviral agent in an amount to treat a bacterial or viral infection. The methods and compositions of the invention are useful for eliciting an immune response against an Epha2-expressing cell, and for treating a hyperproliferative disorder such as cancer (e.g. cancer of the skin, lung, colon, ovary, esophagus, breast, prostate, bladder or pancreas or renal cell carcinoma or melanoma), a non-neoplastic hyperproliferative disorder (e.g. an epithelial cell disorder, specifically asthma, chronic pulmonary obstructive disease, lung fibrosis, bronchial hyperresponsiveness, psoriasis, or seborrheic dermatitis), or a disease involving aberrant angiogenesis (e.g. macular degeneration, diabetic retinopathy, retinopathy of prematurity, vascular restenosis, infantile hemangioma, verruca vulgaris, Kaposi's sarcoma, neurofibromatosis, recessive dystrophic epidermolysis bullosa, rheumatoid arthritis, ankylosing spondylitis, systemic lupus, psoriatic arthropathy, Reiter's syndrome, and Sjogren's syndrome, endometriosis, preeclampsia, atherosclerosis or coronary artery disease). This sequence represents a plasmid containing a multiple cloning site (MCS). It is used in the exemplification of the method of the invention.

Sequence 9808 BP; 2836 A; 1923 C; 1946 G; 3103 T; 0 U; 0 Other;

| | | |
|----------------------------|------|--|
| Query Match | | 89.0%; Score 1116; DB 14; Length 9808; |
| Best Local Similarity | | 98.6%; Pred. No. 0; |
| Matches 1136; Conservative | | 0; Mismatches 15; Indels 1; Gap# 1; |
| QY | 50 | AAACGGATCTATCGTCAATTTACCTCCACGGGAGAGCCTCAGCAAACTGCGCTCA 109 |
| DB | 9192 | ATAAATGATCATATCGTCAATTTACCTCCACGGGAGAGCCTCAGCAAACTGCGCTCA 9133 |
| QY | 110 | GCATTGTAGAGACACGGTCACACTGCTTCGGGTAGTCAATTAACCGGTAAACCA 169 |
| DB | 9132 | GGCATTTGAGAGACACGGTCACACTGCTTCGGGTAGTCAATTAACCGGTAAACCA 9073 |
| QY | 170 | ATAGACATAAGCGGCTATTTAACGACCTGCGCTGAACCGACGCGGTCGAATTTG 229 |
| DB | 9072 | ATAGACATAAGCGGCTATTTAACGACCTGCGCTGAACCGACGCGGTCGAATTTG 9013 |
| QY | 230 | TTGGAATTTCCGCAATTCATCGCTTATATCACTTTATCAGGCGTAGCAACCGCGGT 289 |
| DB | 9012 | TTGGAATTTCCGCAATTCATCGCTTATATCACTTTATCAGGCGTAGC-ACCAGCGGT 8954 |
| QY | 290 | TAAGGCGCACCAATACGCTTAAATAATTAAGCGCGCGCTGCACCTATCGCAGTAC 349 |
| DB | 8953 | TAAGGCGCACCAATACGCTTAAATAATTAAGCGCGCGCTGCACCTATCGCAGTAC 8894 |
| QY | 350 | TCCTGTAAATTCATTAAAGCATTCGCGACATGGAAGCCATCAAAACGCGCATGTAAC 409 |
| DB | 8893 | TGTTGTAAATTCATTAAAGCATTCGCGACATGGAAGCCATCAAAACGCGCATGTAAC 8834 |
| QY | 410 | TGAATCGCGCGGCGCATCAGCACTTGTGCGCTTTCGGTATATATTTGCCCATGTTGAAA 469 |
| DB | 8833 | TGAATCGCGCGGCGCATCAGCACTTGTGCGCTTTCGGTATATATTTGCCCATGTTGAAA 8774 |
| QY | 470 | ACGGGGGCGGAGAGTTGTCATATTTGGCCAGCTTTAAATCAAACTCGTGAACCTACC 529 |
| DB | 8773 | ACGGGGGCGGAGAGTTGTCATATTTGGCCAGCTTTAAATCAAACTCGTGAACCTACC 8714 |

| | | |
|-----------|--|--|
| QY | 530 | CAGGATTTGGCTGAGACGACGAAAAACATATTTCTCAATAAAACCCCTTTAGGGAAATAGCCAGG 589 |
| DB | 8713 | CAGGATTTGGCTGAGACGACGAAAAACATATTTCTCAATAAAACCCCTTTAGGGAAATAGCCAGG 8654 |
| QY | 590 | TTTTACCGTTAACACGCGCACATCTTGGCAATATATGTAGAAAACGCGGAAATCTGTCG 649 |
| DB | 8653 | TTTTACCGTTAACACGCGCACATCTTGGCAATATATGTAGAAAACGCGGAAATCTGTCG 8594 |
| QY | 650 | TGGTATTCACATCCAGAGGCGATGAAAACGTTTCAGTTTCTCATGAAAACGCGTGTAAACA 709 |
| DB | 8593 | TGGTATTCACATCCAGAGGCGATGAAAACGTTTTCAGTTTCTCATGAAAACGCGTGTAAACA 8534 |
| QY | 710 | GGGTGAACACATATCCCATATATCACCAGCTCACCGCTCTTTTCATTGGCCATACGTAATTCGGA 769 |
| DB | 8533 | GGGTGAACACATATCCCATATATCACCAGCTCACCGCTCTTTTCATTGGCCATACGTAATTCGGA 8474 |
| QY | 770 | TGAGCATTCATCAGCGCGGCAAGATGTGAATAAAGCGCGGATAAACCTTGTGCTTATTT 829 |
| DB | 8473 | TGAGCATTCATCAGCGCGGCAAGATGTGAATAAAGCGCGGATAAACCTTGTGCTTATTT 8414 |
| QY | 830 | TTCTTTACGGTCTTTTAAAGCGCGGCAAGATGTGAATAAAGCGCGGATAAACCTTGTGCTTATTT 889 |
| DB | 8413 | TTCTTTACGGTCTTTTAAAGCGCGGCAAGATGTGAATAAAGCGCGGATAAACCTTGTGCTTATTT 8354 |
| QY | 890 | TGAGCAACTGACTGAAATGCTCAAAATGTTTCTTTACGATGCCATTTGGGATATATCAACG 949 |
| DB | 8353 | TGAGCAACTGACTGAAATGCTCAAAATGTTTCTTTACGATGCCATTTGGGATATATCAACG 8294 |
| QY | 950 | GTGTATATTCAGTGAATTTTTTCTCCATTTTAGTCTTCCTTAGCTCTCGAATAATCTCGAC 1009 |
| DB | 8293 | GTGTATATTCAGTGAATTTTTTCTCCATTTTAGTCTTCCTTAGCTCTCGAATAATCTCGAT 8234 |
| QY | 1010 | AACTCAAAAATACGCGGCTAGTGTATTTTCATTTATGTTGTAAGCTTGAACCTCTT 1069 |
| DB | 8233 | AACTCAAAAATACGCGGCTAGTGTATTTTCATTTATGTTGTAAGCTTGAACCTCTT 8174 |
| QY | 1070 | ACGTGCGGATCAACGCTCTCATTTTTCGCAAAAGTTGGCCAGGCTTCCCGGTATCAACA 1129 |
| DB | 8173 | ACGTGCGGATCAACGCTCTCATTTTTCGCAAAAGTTGGCCAGGCTTCCCGGTATCAACA 8114 |
| QY | 1130 | GGGACACAGGATTTATTTATTTTCGGAAGTGTATCTCCGTCACAGGTATTTATTCGTC 1189 |
| DB | 8113 | GGGACACAGGATTTATTTATTTTCGGAAGTGTATCTCCGTCACAGGTATTTATTCGTCG 8054 |
| QY | 1190 | GAAAGGATCCG 1201 |
| DB | 8053 | CAAAGTGGTCG 8042 |
| RESULT 14 | | |
| AAV28398 | | |
| ID | AAV28398 | standard; DNA; 5838 BP. |
| XX | AAV28398; | |
| AC | AAV28398; | |
| XX | 24-JUL-1998 | (first entry) |
| DT | 24-JUL-1998 | (first entry) |
| XX | Plasmid pRZT11 nucleotide sequence. | |
| DE | Tn5 transposase; modified; enzyme; in vitro transposition; mutant; | |
| XX | target; marker; transposon 5; plasmid pRZT11; ds. | |
| KW | Synthetic. | |
| XX | | |
| OS | | |
| FH | Key | Location/Qualifiers |
| FT | insertion_seq | 1..19 |
| FT | | /tag= a |
| FT | CDS | 77..1267 |
| FT | | /tag= b |
| FT | | /note= "tetracycline resistance" |
| FT | | complement (2301..2960) |
| FT | | /tag= c |

| | | | |
|----|--|---|------|
| FT | insertion_seq | /note= "chloramphenicol resistance" | |
| FT | 4564..4582 | /*tag= d | |
| FT | 4715..5530 | /*tag= e | |
| FT | | /note= "kanamycin resistance" | |
| XX | WO9810077-A1. | | |
| PN | | | |
| XX | | | |
| PD | 12-MAR-1998. | | |
| XX | | | |
| PF | 09-SEP-1997; | 97WO-US015941. | |
| XX | | | |
| PR | 09-SEP-1996; | 96US-00814877. | |
| PR | 02-MAY-1997; | 97US-00850880. | |
| XX | | | |
| PA | (WISC) WISCONSIN ALUMNI RES FOUND. | | |
| XX | | | |
| PI | Reznikoff WS, Goryshin IY, Zhou H; | | |
| XX | | | |
| DR | WPI. 1998-193627/17. | | |
| DR | P-PSDB; AAW56695, AAW56696, AAW56697. | | |
| XX | | | |
| PT | Modified Tn5 transposase construct used in novel system for in vitro transposition - used to, e.g. create absolute defective mutants, provide selective markers and to facilitate insertion of specialised DNA sequences into target DNA. | | |
| PT | | | |
| PT | | | |
| PT | | | |
| XX | | | |
| PS | Disclosure; Page 35-46; 73pp; English. | | |
| XX | | | |
| CC | This is the nucleotide sequence of the plasmid pRZTL1 which is used to demonstrate in vitro transposition of a transposable element located between a pair of Tn5 (transposon 5) outside end (OE) termini. The invention provides a genetic construct that contains a nucleotide sequence encoding a modified Tn5 transposase enzyme that has both greater avidity for Tn5 OE repeats and is less likely to assume an inactive multimeric form than a wild type Tn5 transposase and a transposable DNA sequence flanked at its 5' and 3' ends by an 18 or 19 base pair flanking DNA sequence comprising nucleotide A at position 10, T at 11 and A at 12. The modified Tn5 transposase and the transposable DNA which is a DNA donor molecule are used in a system for in vitro transposition. The system and method can be used to create absolute defective mutants, to provide selective markers to target DNA, to provide portable regions of homology to a target DNA, to facilitate insertion of specialised DNA sequences into target DNA, to provide primer binding sites or tags for DNA sequencing, to facilitate production of genetic fusion for gene expression studies and protein domain mapping, as well as to bring together other desired combinations of DNA sequences (combinatorial genetics). The modified Tn5 transposase facilitates in vitro transposition reaction rates of at least about 100-fold higher than can be achieved using wild type transposase (as measure in vivo). In vitro transposition using this system can also use donor DNA and target DNA that is circular or linear. The system also requires no outside high energy source and no other protein other than the modified transposase | | |
| SQ | Sequence 5838 BP; 1370 A; 1525 C; 1516 G; 1427 T; 0 U; 0 Other; | | |
| | Query Match 88.9%; Score 1114.4; DB 2; Length 5838; | | |
| | Best Local Similarity 98.5%; Pred. No. 0; | | |
| | Matches 1135; Conservative 0; Mismatches 16; Indels 1; Gaps 1; | | |
| QY | 50 | AAACCGATCCTATCGTCAATTTATTACCTCCAGGGGAGAGCCTGAGCAACTGGCCTCA | 109 |
| Db | 2033 | ATAAATGATCATATCGTCAATTTATTACCTCCAGGGGAGAGCCTGAGCAACTGGCCTCA | 2092 |
| QY | 110 | GGCATTTGAGAGCAGCACGGTCACACTGCTCCGGTAGTCAATAAACCGGTAACACGCA | 169 |
| Db | 2093 | GGCATTTGAGAGCAGCACGGTCACACTGCTCCGGTAGTCAATAAACCGGTAACACGCA | 2152 |
| QY | 170 | ATAGACATAAGCGGCTATTTAACGACCTGCCCTGAACCGACGACCGGTCGAATTGCT | 229 |
| Db | 2153 | ATAGACATAAGCGGCTATTTAACGACCTGCCCTGAACCGACGACCGGTCGAATTGCT | 2212 |

| | | | |
|----|-------------------------------------|--|------|
| QY | 230 | TTTCGAATTTCTGCCATTCATCCGCTTATATCACTTATTCAGGCGTAGCAACACGAGCGTT | 289 |
| Db | 2213 | TTTCGAATTTCTGCCATTCATCCGCTTATATCAATATTTATCAGGCGTAGC-ACCAGGCGTT | 2271 |
| QY | 290 | TAAAGGCACCAATAACTGCTTTAAAAAAATTTACGCCCGCCCTGCGCACTCATCGCAGTAC | 349 |
| Db | 2272 | TAAAGGCACCAATAACTGCTTTAAAAAAATTTACGCCCGCCCTGCGCACTCATCGCAGTAC | 2331 |
| QY | 350 | TGTTTGAATTCATTAAAGCAITTTCTGCCGACATGGAAGCCATCACAAACGGCATGATGAACC | 409 |
| Db | 2332 | TGTTTGAATTCATTAAAGCAITTTCTCCGACATGGAAGCCATCACAGACGGCATGATGAACC | 2391 |
| QY | 410 | TGAATCGCCAGCGGCATCAGCACCTTGTGCGCTTGCGTATAATTTAGCCCATGTTGAAA | 469 |
| Db | 2392 | TGAATCGCCAGCGGCATCAGCACCTTGTGCGCTTGCGTATAATTTAGCCCATGTTGAAA | 2451 |
| QY | 470 | ACGGGGGCGAAGAGTTGCTCCATATTGGCCACGTTTAAATCAAACTGGTGAACCTCAC | 529 |
| Db | 2452 | ACGGGGGCGAAGAGTTGCTCCATATTGGCCACGTTTAAATCAAACTGGTGAACCTCAC | 2511 |
| QY | 530 | CAGGGATTGGCTGAGACGAAAAACAATTTCTCAATAAAACCTTTAGGGAATATAGGCCAGG | 589 |
| Db | 2512 | CAGGGATTGGCTGAGACGAAAAACAATTTCTCAATAAAACCTTTAGGGAATATAGGCCAGG | 2571 |
| QY | 590 | TTTTTACCCTAACCGCCACATCTTGGGAATATATGTGTAGAAACTCCCGAAATCGTCG | 649 |
| Db | 2572 | TTTTTACCCTAACCGCCACATCTTGGGAATATATGTGTAGAAACTCCCGAAATCGTCG | 2631 |
| QY | 650 | TGGTATTCACTCCAGAGCGATGAAAAAGTTTCAGTTTGCTCATGGAAAAAGGTGTACAA | 709 |
| Db | 2632 | TGGTATTCACTCCAGAGCGATGAAAAAGTTTCAGTTTGCTCATGGAAAAAGGTGTACAA | 2691 |
| QY | 710 | GGTGCAACACTATCCCATATCACAGACTCACCGCTTTTCATGTGCGCATACGTAAATCCGGA | 769 |
| Db | 2692 | GGTGCAACACTATCCCATATCACAGACTCACCGCTTTTCATGTGCGCATACGTAAATCCGGA | 2751 |
| QY | 770 | TCAGCATTCATCAGCGCGGCAAGAAATGTGAATAAAGCCGGATAAAACTTGTGCTTATTT | 829 |
| Db | 2752 | TCAGCATTCATCAGCGCGGCAAGAAATGTGAATAAAGCCGGATAAAACTTGTGCTTATTT | 2811 |
| QY | 830 | TTCTTTAGGTCCTTTAAAAAGGCGGTAATATCCAGCTGAAACGGTCTCGTTATAGGTACAT | 889 |
| Db | 2812 | TTCTTTAGGTCCTTTAAAAAGGCGGTAATATCCAGCTGAAACGGTCTCGTTATAGGTACAT | 2871 |
| QY | 890 | TCAGCAACTGACTGAATAGCCTCAAAATGTTCTTTACGATGCCATTGGGATATATCAACG | 949 |
| Db | 2872 | TCAGCAACTGACTGAATAGCCTCAAAATGTTCTTTACGATGCCATTGGGATATATCAACG | 2931 |
| QY | 950 | GTGGGTATATCCAGTGATTTTTTTTCTCCATTTTACGCTTCCTTAGCTCCTGAAATCTCGAC | 1009 |
| Db | 2932 | GTGGGTATATCCAGTGATTTTTTTTCTCCATTTTACGCTTCCTTAGCTCCTGAAATCTCGAT | 2991 |
| QY | 1010 | AACCTCAAAATAATACGCCCGGTAGTGATCTTATTTTCATTTATGTTGAAAGTTGGAACTCTTT | 1069 |
| Db | 2992 | AACCTCAAAATAATACGCCCGGTAGTGATCTTATTTTCATTTATGTTGAAAGTTGGAACTCTTT | 3051 |
| QY | 1070 | ACGTGCCGATCAACGCTCTCATTTTTCGCCAAAAGTTGGCCCGAGGCTTCCCGGTATCAACA | 1129 |
| Db | 3052 | ACGTGCCGATCAACGCTCTCATTTTTCGCCAAAAGTTGGCCCGAGGCTTCCCGGTATCAACA | 3111 |
| QY | 1130 | GGGACACACAGGATTTATTTATTTCTGCGAAGTGATCTTCCGTCACAGGTATTTATTCGGTC | 1189 |
| Db | 3112 | GGGACACACAGGATTTATTTATTTCTGCGAAGTGATCTTCCGTCACAGGTATTTATTCGGTC | 3171 |
| QY | 1190 | GAAGAAGGATCCG 1201 | |
| Db | 3172 | CAAGTGGGTGCG 3183 | |
| | RESULT 15 | | |
| | AA206434 | | |
| | ID AAZ06434 standard; DNA; 5838 BP. | | |
| | XX | | |

RESULT 15
AAZ06434
ID AAZ06434 standard; DNA; 5838 BP.
XX

AC AAZ06434;
 XX 09-NOV-1999 (first entry)
 XX Plasmid pRZTL1, target DNA for Tn5 transposase assay.
 XX DE transposase; modified form; wildtype; multimeric; OE termini; IE termini;
 XX KW outside end termini; inside end termini; plasmid; repeat sequence;
 XX KW mutation; cyclic; ds.
 XX OS Synthetic.
 XX PH Key Location/Qualifiers
 FT insertion_seq 1..19
 FT /*tag= a
 FT CDS 77..1267
 FT /*tag= b
 FT /gene= "Tetr"
 FT /function= "effluxes tet"
 FT /product= "inner membrane protein"
 FT /phenotype= "Tetracycline resistance"
 FT CDS 2301..2960
 FT /*tag= c
 FT /gene= "ChlR"
 FT /function= "acetylates chl"
 FT /product= "chloramphenicol acetyltransferase"
 FT /function= "acetylates chl"
 FT /phenotype= "chloramphenicol resistance"
 FT insertion_seq 4564..4582
 FT /*tag= d
 FT CDS 4715..5530
 FT /*tag= e
 FT /function= "blocks action of kanamycin"
 FT /phenotype= "kanamycin resistance"
 XX US5948622-A.
 XX 07-SEP-1999.
 XX 06-OCT-1997; 97US-00944916.
 XX 09-SEP-1996; 96US-00814877.
 XX 02-MAY-1997; 97US-00850880.
 XX (WISC) WISCONSIN ALUMNI RES FOUND.
 XX Zhou H, York DL, Goryshin IY, Reznikoff WS;
 XX WPI; 1999-517947/43.
 XX P-PSDB; AAV15383, AAV15384, AAV15385.
 XX In vitro transposition using a Tn5 based genetic construct.
 XX Example 1; Col 31-39; 48pp; English.
 XX This plasmid can be used in the assay of transposition frequency by the
 CC modified Tn5 transposase. The insertion of the Tn5 transposase releases a
 CC fragment of the plasmid that corresponds to the region of Kanamycin
 CC resistance gene that lies between the two OE sequences. The plasmid was
 CC also used to investigate the nature of the termini involved in the
 CC transposition reaction. Wildtype Outside End (OE,AAZ06435) and Inside End
 CC (IE, AAZ06438) were compared and an effort made to randomise the
 CC nucleotides at each of the seven positions of difference. This eventually
 CC lead to the identification of the mutant OE sequences AAZ06436 and
 CC AAZ06437
 XX
 XX Query Match 88.9%; Score 1114.4; DB 2; Length 5838;
 XX Best Local Similarity 98.5%; Pred. No. 0;
 XX Matches 1135; Conservative 0; Mismatches 16; Indels 1; Gaps 1;
 XX 50 AAAACGGATCCTATCGTCAATTATTACCTCCAGGGGAGAGCGCTGAGCAAACTGCGCTCA 109

2033 ATAAATGATCATATCGTCAATTTATTACCTCCAGGGGAGAGCTGAGCAAACTGCGCTCA 2092
 110 GGCAATTTGAGAAGACACAGGTGCTTCCGCTAGTCAATAAAACCGGTAAACACGACA 169
 2093 GGCAATTTGAGAAGACACAGGTGCTTCCGCTAGTCAATAAAACCGGTAAACACGACA 2152
 170 ATAGACATAAGCGGCTATTAAACGACCTGCGCTGAAACGACGCGGTGCGAATTTGCT 229
 2153 ATAGACATAAGCGGCTATTAAACGACCTGCGCTGAAACGACGCGGTGCGAATTTGCT 2212
 230 TTCGAATTTCTGCCAATTCATCGCTTATATCACTTATTCAGGCGTAGCAAAACGAGCGTT 289
 2213 TTCGAATTTCTGCCAATTCATCGCTTATATCACTTATTCAGGCGTAGC-ACCAGGCGTT 2271
 290 TAAGGGCACCAATAACTGCTTTAAAAAATAATACGCGCCGCTGCGCACTCATCGCAGTAC 349
 2272 TAAGGGCACCAATAACTGCTTTAAAAAATAATACGCGCCGCTGCGCACTCATCGCAGTAC 2331
 350 TGTGTGAATTCATTAAAGCAATTCGCGACATGGAAGCCATCAAAACGCGCATGATGAACC 409
 2332 TGTGTGAATTCATTAAAGCAATTCGCGACATGGAAGCCATCAAAACGCGCATGATGAACC 2391
 410 TGAATCGCCAGCGGCATCAGCACCTTGTGCGCTTGGGTATATATTTGCCCATGTTGAAA 469
 2392 TGAATCGCCAGCGGCATCAGCACCTTGTGCGCTTGGGTATATATTTGCCCATGTTGAAA 2451
 470 ACGGGGGCGAAGAAGTTGTCCATATTTGCCACGTTTAAATCAAACTGGTGAACCTCACC 529
 2452 ACGGGGGCGAAGAAGTTGTCCATATTTGCCACGTTTAAATCAAACTGGTGAACCTCACC 2511
 530 CAGGATTTGGCTGAGACGAAAAACAATTTCTCAATTAACCCCTTTAGGGAAAAATAGCCAGG 589
 2512 CAGGATTTGGCTGAGACGAAAAACAATTTCTCAATTAACCCCTTTAGGGAAAAATAGCCAGG 2571
 590 TTTTCACCGTAAACACGCCACATCTTGGCAATATATGTAGAAACTGCGGAAATCGTCG 649
 2572 TTTTCACCGTAAACACGCCACATCTTGGCAATATATGTAGAAACTGCGGAAATCGTCG 2631
 650 TGGTATTTCACTCCAGAGCGATGAAAAAGTTTCAGTTTGTCTCATGAAAAACGGTGTAAACAA 709
 2632 TGGTATTTCACTCCAGAGCGATGAAAAAGTTTTCAGTTTGTCTCATGAAAAACGGTGTAAACAA 2691
 710 GGGTGAACACATATCCCATATCACAGCTCACCGTCTTTTCATTGGCCATACGTAATTTCCGGA 769
 2692 GGGTGAACACATATCCCATATCACAGCTCACCGTCTTTTCATTGGCCATACGTAATTTCCGGA 2751
 770 TGAGCAATTCATCAGCGGGCGAAGATGTGAATAAAGCGCGGATATAAACTTGTGCTTATTT 829
 2752 TGAGCAATTCATCAGCGGGCGAAGATGTGAATAAAGCGCGGATATAAACTTGTGCTTATTT 2811
 830 TTTCTTTAGGCTTTTAAAAAGCGCGTATATCCAGCTGNAACGGTCTGGTTATAGGTACAT 889
 2812 TTTCTTTAGGCTTTTAAAAAGCGCGTATATCCAGCTGNAACGGTCTGGTTATAGGTACAT 2871
 890 TGAGCAACTGACTGAAATGCGCTCAAAATGTTCTTTTACGATGCCATTGGGATATATCAACG 949
 2872 TGAGCAACTGACTGAAATGCGCTCAAAATGTTCTTTTACGATGCCATTGGGATATATCAACG 2931
 950 GTGGTATATCCAGTGAATTTTTCCTCAATTTTAGTCTTCTTAGCTTCTGAAAAATCTCGAC 1009
 2932 GTGGTATATCCAGTGAATTTTTCCTCAATTTTAGTCTTCTTAGCTTCTGAAAAATCTCGAT 2991
 1010 AACTCAAAAAATACGCGCGGTAGTATCTTATTTTCAATTTAGTGAAGTGTGAACCTCTT 1069
 2992 AACTCAAAAAATACGCGCGGTAGTATCTTATTTTCAATTTAGTGAAGTGTGAACCTCTT 3051
 1070 ACGTCCCGATCAACGCTCTCATTTTTCGCAAAAGTTGGCCCGGCTTCCCGGTATCAACA 1129
 3052 ACGTCCCGATCAACGCTCTCATTTTTCGCAAAAGTTGGCCCGGCTTCCCGGTATCAACA 3111
 1130 GGGACACAGGATTTATTTATTTCTCGGAAGTGAATCTTCCGTCACAGGTATTTATTTCCGTC 1189

Db 3112 GGGACACACAGGATTTATTCTCGGAAGTGATCTTCGTCCACAGGTATTTATTTCGGCG 3171
Qy 1190 GAAAGGATCCG 1201
Db 3172 CAAAGTGGCTCG 3183

Search completed: January 17, 2006, 19:35:10
Job time : 784.788 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2006 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 17, 2006, 18:09:28 ; Search time 5307.56 Seconds
(without alignments)
11054.239 Million cell updates/sec

Title: US-10-511-327-2
Perfect score: 1254
Sequence: 1 gatctgattgattgaacgaa.....ttcgttcaatcaatcagatc 1254

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0
Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650
Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_hic:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_est7:*
9: gb_gss1:*
10: gb_gss2:*
11: gb_gss3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match % | Length | DB | ID | Description |
|------------|-------|---------------|--------|----|----------|--------------------|
| 1 | 867.4 | 69.2 | 971 | 10 | CL111359 | CL111359 ISB1-55D1 |
| 2 | 858.8 | 68.5 | 1039 | 10 | CL094006 | CL094006 ISB1-24P6 |
| 3 | 854.6 | 68.1 | 973 | 10 | CL112191 | CL112191 ISB1-56K1 |
| 4 | 852 | 67.9 | 968 | 10 | CL131987 | CL131987 ISB1-101G |
| 5 | 852 | 67.9 | 994 | 10 | CL134875 | CL134875 ISB1-105K |
| 6 | 842.4 | 67.2 | 1018 | 10 | CL132765 | CL132765 ISB1-102I |
| 7 | 838 | 66.8 | 1047 | 10 | CL091480 | CL091480 ISB1-102I |
| 8 | 832.2 | 66.4 | 1001 | 10 | CL145346 | CL145346 ISB1-145L |
| 9 | 831.4 | 66.3 | 971 | 10 | CL129946 | CL129946 ISB1-97H8 |
| 10 | 829.8 | 66.2 | 968 | 10 | CL131806 | CL131806 ISB1-101C |
| 11 | 824.6 | 65.8 | 1007 | 10 | CL112441 | CL112441 ISB1-57C1 |
| 12 | 815.8 | 65.1 | 945 | 11 | CT015391 | CT015391 KBRH124J0 |
| 13 | 814 | 64.9 | 923 | 10 | CL132302 | CL132302 ISB1-101N |
| 14 | 812.2 | 64.8 | 952 | 10 | CZ941767 | CZ941767 311863 To |
| 15 | 803.6 | 64.1 | 901 | 10 | CL129968 | CL129968 ISB1-97H2 |
| 16 | 798.4 | 63.7 | 972 | 10 | CL091218 | CL091218 ISB1-20B5 |
| 17 | 797.4 | 63.6 | 939 | 9 | CC219264 | CC219264 CH261-114 |
| 18 | 792.6 | 63.2 | 1001 | 10 | CL120945 | CL120945 ISB1-79L1 |
| 19 | 783.2 | 62.5 | 1052 | 10 | CL114391 | CL114391 ISB1-60C1 |
| 20 | 781.4 | 62.3 | 1061 | 10 | CL113434 | CL113434 ISB1-58I1 |
| 21 | 780.8 | 62.3 | 831 | 9 | BH239118 | BH239118 ATYCK59TF |
| 22 | 772 | 61.6 | 1078 | 10 | CL110921 | CL110921 ISB1-54H1 |

| | | | | | | |
|----|-------|------|------|----|----------|--------------------|
| 23 | 769.4 | 61.4 | 850 | 8 | CV979827 | CV979827 UMC-bemiv |
| 24 | 769.2 | 61.3 | 1007 | 10 | CL114392 | CL114392 ISB1-60C1 |
| 25 | 767 | 61.2 | 965 | 10 | CL134505 | CL134505 ISB1-105B |
| 26 | 765.8 | 61.1 | 812 | 10 | DU053555 | DU053555 141529 To |
| 27 | 765.4 | 61.0 | 984 | 10 | CL132776 | CL132776 ISB1-102I |
| 28 | 762 | 60.8 | 1004 | 10 | CL039461 | CL039461 CH216-470 |
| 29 | 760.6 | 60.7 | 1011 | 10 | CL111697 | CL111697 ISB1-55K1 |
| 30 | 760.4 | 60.6 | 1026 | 10 | CL114498 | CL114498 ISB1-60F2 |
| 31 | 759.2 | 60.5 | 765 | 10 | CL133473 | CL133473 ISB1-103I |
| 32 | 758.8 | 60.5 | 955 | 10 | CL144387 | CL144387 ISB1-144G |
| 33 | 756.8 | 60.4 | 1207 | 9 | CC279944 | CC279944 CH261-24C |
| 34 | 756.6 | 60.3 | 1027 | 10 | CL111508 | CL111508 ISB1-55G8 |
| 35 | 756.4 | 60.3 | 794 | 9 | BH243026 | BH243026 AUTG06TR |
| 36 | 756.2 | 60.3 | 1010 | 10 | CL112625 | CL112625 ISB1-57G1 |
| 37 | 754.4 | 60.2 | 989 | 10 | CL132140 | CL132140 ISB1-101K |
| 38 | 752.4 | 60.0 | 969 | 10 | CL124573 | CL124573 ISB1-85K6 |
| 39 | 750.2 | 59.8 | 782 | 9 | BH242215 | BH242215 ATZPD43TF |
| 40 | 750.2 | 59.8 | 801 | 10 | DU058617 | DU058617 90433 Tom |
| 41 | 750 | 59.8 | 934 | 10 | CL124145 | CL124145 ISB1-85A9 |
| 42 | 749.2 | 59.7 | 939 | 10 | CL132916 | CL132916 ISB1-102L |
| 43 | 748.6 | 59.7 | 978 | 10 | CL033185 | CL033185 CH216-36C |
| 44 | 748.2 | 59.7 | 985 | 10 | CL131971 | CL131971 ISB1-101G |
| 45 | 746.8 | 59.6 | 946 | 10 | CL145878 | CL145878 ISB1-146H |

ALIGNMENTS

RESULT 1
CL111359
LOCUS ISB1-55D1.T7.1 ISB1 Xenopus tropicalis genomic clone ISB1-55D1, linear GSS 05-JAN-2004
DEFINITION genomic survey sequence.
ACCESSION CL111359
VERSION CL111359.1 GI:40604994
KEYWORDS GSS.
ORGANISM Xenopus tropicalis (western clawed frog)
SOURCE Xenopus tropicalis
Xenopus tropicalis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
Xenopodinae; Xenopus; Silurana.
REFERENCE 1 (Bases 1 to 971)
Kremitzki,C., Carter,J., McPherson,J., Warren,W., Graves,T.,
Mardis,E. and Wilson,R.
TITLE A physical map of the xenopus tropicalis genome
JOURNAL Unpublished (2003)
COMMENT Contact: Richard K Wilson
Genome Sequencing Center
Washington University School of Medicine
Email: submissions@watson.wustl.edu
Insert Length: 75000 Std Error: 0.00
Seq primer: T7 TAATACGACTCATATAGGG
Class: BAC ends
High quality sequence start: 8
High quality sequence stop: 824.
Location/Qualifiers
1. .971
/organism="Xenopus tropicalis"
/mol_type="genomic DNA"
/db_xref="taxon:8364"
/clone="ISB1-55D1"
/clone_lib="ISB1"
/note="Vector: pBelOBAC11; ISB-1 Xenopus tropicalis BAC
Library Segment 1"

ORIGIN
Query Match 69.2%; Score 867.4; DB 10; Length 971;
Best Local Similarity 99.1%; Pred. No. 8.6e-254;
Matches 882; Conservative 0; Mismatches 7; Indels 1; Gaps 1;
Qy 231 TCGAATTTCTGCCATTATTCGCTTATTTATCATTATTCAGCGTAGCAACGCGTTT 290
|||||
Db 83 TCGAATTTCTGCCATTATTCGCTTATTTATCATTATTCAGCGTAGCAACGCGTTT 142
|||||

QY 291 AAGGCGCAATAAATGCTTAAATAAATAGCGCCCGCCCTGCGACATCATCGCAGTACT 350
 DB 143 AAGGCGCAATAAATGCTTAAATAAATAGCGCCCGCCCTGCGACATCATCGCAGTACT 202
 QY 351 GTTGTAAATTCATTAAGCAATTCGCGGACATGGAAGCCATCAAAACGGCATGATGAACCT 410
 DB 203 GTTGTAAATTCATTAAGCAATTCGCGGACATGGAAGCCATCAAAACGGCATGATGAACCT 262
 QY 411 GAATCGCGAGGCGCATCAGCACCTTTGTCGCTTGGTATTAATTAATTTGCCCATGTTGAAAA 470
 DB 263 GAATCGCGAGGCGCATCAGCACCTTTGTCGCTTGGTATTAATTAATTTGCCCATGTTGAAAA 322
 QY 471 CGGCGCGGAGAAAGTGTTCATTAATTTGGGCAAGCTTTAAATCAAACTGTGTGAACCTCACCC 530
 DB 323 CGGCGCGGAGAAAGTGTTCATTAATTTGGGCAAGCTTTAAATCAAACTGTGTGAACCTCACCC 382
 QY 531 AGGGATTGGCTGAGACGAAACATATCTCAATAAACCTTTAGGGAATAGGCCAGGT 590
 DB 383 AGGGATTGGCTGAGACGAAACATATCTCAATAAACCTTTAGGGAATAGGCCAGGT 442
 QY 591 TTTCACCGGTAAACCGCCACATCTTCGGAATATATGTAGAACTGCGCGAAATCGTCGT 650
 DB 443 TTTCACCGGTAAACCGCCACATCTTCGGAATATATGTAGAACTGCGCGAAATCGTCGT 502
 QY 651 GGATATTCATCCAGAGCGATGAAACGTTTCAGTTTGTCTCATGTGAAACGGTGTAAACAG 710
 DB 503 GGATATTCATCCAGAGCGATGAAACGTTTCAGTTTGTCTCATGTGAAACGGTGTAAACAG 562
 QY 711 GGTGAACATATCCCATATCCAGCTCACCGTCTTTTCAATTTGCCATAGTAATTCGGAT 770
 DB 563 GGTGAACATATCCCATATCCAGCTCACCGTCTTTTCAATTTGCCATAGTAATTCGGAT 622
 QY 771 GAGCATTCATCAGCGGGGCAAGAATGTGAATAAGCGCGGATATAAATCTGTCTTATTTT 830
 DB 623 GAGCATTCATCAGCGGGGCAAGAATGTGAATAAGCGCGGATATAAATCTGTCTTATTTT 682
 QY 831 TCTTTACGGTCTTTAAAGGCGGTAATATCCAGCTGAAACGGTCTGTGTATAGGTACAT 890
 DB 683 TCTTTACGGTCTTTAAAGGCGGTAATATCCAGCTGAAACGGTCTGTGTATAGGTACAT 742
 QY 891 GAGCACTACGTAAGTCTCAAAATGTTCTTACGATGCAACGGTCTGTGTATAGGTACAT 950
 DB 743 GAGCACTACGTAAGTCTCAAAATGTTCTTACGATGCAACGGTCTGTGTATAGGTACAT 802
 QY 951 TGGTATATCCAGTATTTTCTCCATTTTACGTTTCTTACGTTTCTTACGTTTCTTACGTT 1010
 DB 803 TGGTATATCCAGTATTTTCTCCATTTTACGTTTCTTACGTTTCTTACGTTTCTTACGTT 862
 QY 1011 ACTCAAAAATACGCGCGTAGTATCTTATTTTATTTATTTATTTATTTATTTATTTATTT 1070
 DB 863 ACTC-AAAAATACGCGCGTAGTATCTTATTTTATTTATTTATTTATTTATTTATTTATTT 921
 QY 1071 CGTGGCGATCAAGTCTCATTTTCGCAAAAGTTGCGCCAGGCTTCCCG 1120
 DB 922 CGGCGCGATTAAGTCTCAATTTTCGCAAAAGTTGCGCCAGGCTTCCCG 971

RESULT 2
 CL094006
 LOCUS
 DEFINITION
 ISBI-24F6_r7.1 ISBI Xenopus tropicalis genomic clone ISBI-24F6,
 genomic survey sequence.
 CL094006
 CL094006.1 GI:40587641
 GSS.
 Xenopus tropicalis (western clawed frog)
 Xenopus tropicalis
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
 Xenopodinae; Xenopus; Silurana.
 1 (bases 1 to 1039)
 REFERENCE
 AUTHORS
 Kremitzki, C., Carter, J., McPherson, J., Warren, W., Graves, T.,

Mardis, E. and Wilson, R.
 A physical map of the xenopus tropicalis genome
 Unpublished (2003)
 Contact: Richard K Wilson
 Genome Sequencing Center
 Washington University School of Medicine
 Email: submissions@wustl.edu
 Insert Length: 75000 Std Error: 0.00
 Seq primer: T7 TAATACGACTCACTATAGG
 Class: BAC ends
 High quality sequence start: 2
 High quality sequence stop: 744.
 FEATURES
 Location/Qualifiers
 source
 1..1039
 /organism="Xenopus tropicalis"
 /mol_type="genomic DNA"
 /db_xref="taxon:8364"
 /clone="ISBI-24F6"
 /clone_lib="ISBI"
 /note="Vector: pBelobAC11; ISB-1 Xenopus tropicalis BAC
 library Segment 1"

ORIGIN
 Query Match 68.5%; Score 858.8; DB 10; Length 1039;
 Best Local Similarity 96.4%; Pred. No. 3.8e-251;
 Matches 889; Conservative 0; Mismatches 32; Indels 1; Gaps 1;
 QY 231 TCGAATTTCTGCCATTCATCCGCTTATATCACTTTTTCAGGCGTAGCAACGCGCTTT 290
 DB 67 TCGAATTTCTGCCATTCATCCGCTTATATCACTTTTTCAGGCGTAGCAACGCGCTTT 126
 QY 291 AAGGCGCACATTAATCGCTTAAATAAATAGCGCCCGCCCTGCGACATCATCGCAGTACT 350
 DB 127 AAGGCGCACATTAATCGCTTAAATAAATAGCGCCCGCCCTGCGACATCATCGCAGTACT 186
 QY 351 GTTGTAAATTCATTAAGCAATTCGCGGACATGGAAGCCATCAAAACGGCATGATGAACCT 410
 DB 187 GTTGTAAATTCATTAAGCAATTCGCGGACATGGAAGCCATCAAAACGGCATGATGAACCT 246
 QY 411 GAATCGCGAGGCGCATCAGCACTTTGTCGCTTGGTATTAATTTGCCCATGTTGAAAA 470
 DB 247 GAATCGCGAGGCGCATCAGCACTTTGTCGCTTGGTATTAATTTGCCCATGTTGAAAA 306
 QY 471 CGGCGCGCAAGAGTTGTCATATTTGGCCACGTTTAAATCAAACTGTTGAAACCTCACCC 530
 DB 307 CGGCGCGCAAGAGTTGTCATATTTGGCCACGTTTAAATCAAACTGTTGAAACCTCACCC 366
 QY 531 AGGGATTGGCTGAGACGAAACATATTTCTCAATAAACCTTTAGGGAATAGGCCAGGT 590
 DB 367 AGGGATTGGCTGAGACGAAACATATTTCTCAATAAACCTTTAGGGAATAGGCCAGGT 426
 QY 591 TTTTACCGTAAACGCGCCATCTTTCGGAATATATGTTAGAACTGCGCGGAATCGTCGT 650
 DB 427 TTTTACCGTAAACGCGCCATCTTTCGGAATATATGTTAGAACTGCGCGGAATCGTCGT 486
 QY 651 GGTATTTCACTCCAGAGCGATGAACGTTTTCAGTTTGTCTCATGGAAGCGTGTAAACAG 710
 DB 487 GGTATTTCACTCCAGAGCGATGAACGTTTTCAGTTTGTCTCATGGAAGCGTGTAAACAG 546
 QY 711 GGTGAACATATCCCATATTCACGCTCACCGTCTTTTCAATTTGCCATATCGTAATTCGGAT 770
 DB 547 GGTGAACATATCCCATATTCACGCTCACCGTCTTTTCAATTTGCCATATCGTAATTCGGAT 606
 QY 771 GAGCATTCATCAGCGCGGCAAGATGTGAATAAGCGCGGATATAAATCTGTCTTATTTT 830
 DB 607 GAGCATTCATCAGCGCGGCAAGATGTGAATAAGCGCGGATATAAATCTGTCTTATTTT 666
 QY 831 TCTTTACCGTCTTTAAAGGCGGTAATATCCAGCTGAAACGGTGTATAGGTACATTT 890
 DB 667 TCTTTACCGTCTTTAAAGGCGGTAATATCCAGCTGAAACGGTGTATAGGTACATTT 726
 QY 891 GAGCAATCTGACTGAAATGCTTCAAAATGTTCTTTACGATGCGATTTGGGATATATCAACCG 950


```

Db      727 GAGCAACTGACTGAATGCTTCAAAATGTTCTTTAGGATGCCATTGGGATATATCAACGG 786
Qy      951 TGGTATATCCAGTGATTTTTTCTCATTTTACGTTCTTACGCTCCTGAAATCTCGACA 1010
Db      787 TGGTATATCCAGTGATTTTTTCTCCATTTTACGTTCTTACGCTCCTGAAATCTCCATA 846
Qy      1011 ACTCAAAAATAACGCCCGGTAGTGATCTTATTTCAATATGGTGAAGTTGGAACTCTTA 1070
Db      847 ACTC-AAAAATAACCCCGGTAGTGATCTTTTTTATATGGTGAAGTTGGAACTCTTA 905
Qy      1071 CGTGGCGATCAACGCTCTCATTTTTCGCAAAAGTTGGCCAGGCTTCCCGGTATCAACAG 1130
Db      906 CGTGGCCCATCAACGCTCTATTTTCCCAAAAGTTGGCCAGGCTTCCCGGTATCAACGG 965
Qy      1131 GACACACAGGATTTATTTATTC 1152
Db      966 GACACCGGATTTCTTTCTC 987

RESULT 3
CL112191
LOCUS      ISB1-56K13 T7.1 ISB1 xenopus tropicalis genomic clone ISB1-56K13,
DEFINITION genomic survey sequence.
ACCESSION CL112191
VERSION    CL112191.1 GI:40605826
KEYWORDS   GSS.
SOURCE     xenopus tropicalis (western clawed frog)
ORGANISM   xenopus tropicalis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
            Xenopodinae; Xenopus; Silurana.
REFERENCE  1 (bases 1 to 973)
            Kremitski, C., Carter, J., McPherson, J., Warren, W., Graves, T.,
            Mardis, E. and Wilson, R.
            A physical map of the xenopus tropicalis genome
            Unpublished (2003)
            Contact: Richard K Wilson
            Genome Sequencing Center
            Washington University School of Medicine
            Email: submissions@watson.wustl.edu
            Insert Length: 75000 Std Error: 0.00
            Seq primer: T7 TAATACGACTCACTATAGG
            Class: BAC ends
            High quality sequence start: 2
            High quality sequence stop: 769.
            Location/Qualifiers
                1..973
                /organism="Xenopus tropicalis"
                /mol_type="genomic DNA"
                /db_xref="taxon:8364"
                /clone="ISB1-56K13"
                /clone_lib="ISB1"
                /note="Vector: pBelobAC11; ISB-1 Xenopus tropicalis BAC
                Library Segment 1"

FEATURES             source
    Query Match      68.1%; Score 854.6; DB 10; Length 973;
    Best Local Similarity 98.9%; Pred. No. 7.2e-250;
    Matches 871; Conservative 0; Mismatches 9; Indels 1; Gaps 1;

Qy      307 GCCTTAAAAAATTACGCCCGCTGCCACTCATCGCAGTACTGTTGTAATTCATTAAAG 366
Db      2 GCCTTAAAAAATTACGCCCGCTGCCACTCATCGCAGTACTGTTGTAATTCATTAAAG 61
Qy      367 CATTCGCGCATGGAAGCCATCAAAACGGCATGATGAACCTGAATCGCCAGGCGAT 426
Db      62 CATTCGCGCATGGAAGCCATCAAAACGGCATGATGAACCTGAATCGCCAGGCGAT 121
Qy      427 CAGCACCTTTCGCTTCGCTGATATATTTGCCCATGTTGAAACGGGGGGAAGAAGTT 486
Db      122 CAGCACCTTTCGCTTCGCTGATATATTTGCCCATGTTGAAACGGGGGGAAGAAGTT 181

```

```

Qy      487 GTCCATATTGGCCACGTTTAAATCAAACTCGTGAACTCACCCAGGGATTGGCTGAGAC 546
Db      182 GTCCATATTGGCCACGTTTAAATCAAACTCGTGAACTCACCCAGGGATTGGCTGAGAC 241
Qy      547 GAAAAACATATTTCTCAATAAAACCCCTTTAGGAAATAGGCCAGGTTTTTCACCGTAACACGC 606
Db      242 GAAAAACATATTTCTCAATAAAACCCCTTTAGGAAATAGGCCAGGTTTTTCACCGTAACACGC 301
Qy      607 CACATTTGCGAATATATGTGTAGAACTCGCGGAAATCGTCTGGTGTATTCATCTCCAGAG 666
Db      302 CACATTTGCGAATATATGTGTAGAACTCGCGGAAATCGTCTGGTGTATTCATCTCCAGAG 361
Qy      667 CGATGAAACCGTTTCAGTTTCCTCATGGAACCGGTGTAAACAGGGTGAACATATCCCA 726
Db      362 CGATGAAACCGTTTCAGTTTCCTCATGGAACCGGTGTAAACAGGGTGAACATATCCCA 421
Qy      727 TATCACCAGCTCACCGTCTTTTCATTGCCATACGTAATTTCCGGATGAGCATTCATCAGCGG 786
Db      422 TATCACCAGCTCACCGTCTTTTCATTGCCATACGGAATTTCCGGATGAGCATTCATCAGCGG 481
Qy      787 GGCAAGATGTGAATAAAGCGCGGATAAACTTGTGCTTATTTTCTTTACGGTCTTTAA 846
Db      482 GGCAAGATGTGAATAAAGCGCGGATAAACTTGTGCTTATTTTCTTTACGGTCTTTAA 541
Qy      847 AAAGCCCGTATATCCAGCTCAACGGTCTGTTTATAGTATAGTATGAGCAACTGACTGAAA 906
Db      542 AAAGCCCGTATATATCCAGCTCAACGGTCTGTTTATAGTATAGTATGAGCAACTGACTGAAA 601
Qy      907 TGCCTCAAAATGTTCTTTTACGATGCCATTGGGATATATCAACGGTGGTATATCCAGTGTAT 966
Db      602 TGCCTCAAAATGTTCTTTTACGATGCCATTGGGATATATCAACGGTGGTATATCCAGTGTAT 661
Qy      967 TTTTCTTCCATTTTAGCTTCTTACGCTCTGAAATCTCGACAACCTCAAAAAATACGCC 1026
Db      662 TTTTCTTCCATTTTAGCTTCTTACGCTCTGAAATCTCGATAACTCAAAAAATACGCC 721
Qy      1027 CGGTAGTGTATCTTATTTCAATATGTTGAAAGTTGGAACCTCTTACGTGCCATCAACGTC 1086
Db      722 CGGTAGTGTATCTTATTTCAATATGTTGAAAGTTGGAACCTCTTACGTGCCATCAACGTC 781
Qy      1087 TCATTTTCGCCCAAAAGTTGGCCAGGCTTCCCGTATCAACAGGGACACACAGGA-TTTTA 1145
Db      782 TCATTTTCGCCCAAAAGTTGGCCAGGCTTCCCGTATCAACAGGGACACACAGGA-TTTTA 841
Qy      1146 TTTATTTCTGGAAGTGTATTTCCGTCACAGGTATTTATTCG 1186
Db      842 TTTATTTCTGGAAGTGTATTTCCGTCGCCGTTATTTATTCG 882

RESULT 4
CL131987
LOCUS      ISB1-101G15 T7.1 ISB1 Xenopus tropicalis genomic clone ISB1-101G15,
DEFINITION genomic survey sequence.
ACCESSION CL131987
VERSION    CL131987.1 GI:40625622
KEYWORDS   GSS.
SOURCE     Xenopus tropicalis (western clawed frog)
ORGANISM   Xenopus tropicalis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
            Xenopodinae; Xenopus; Silurana.
            1 (bases 1 to 968)
            Kremitski, C., Carter, J., McPherson, J., Warren, W., Graves, T.,
            Mardis, E. and Wilson, R.
            A physical map of the xenopus tropicalis genome
            Unpublished (2003)
            Contact: Richard K Wilson
            Genome Sequencing Center
            Washington University School of Medicine
            Email: submissions@watson.wustl.edu
            Insert Length: 75000 Std Error: 0.00
            Seq primer: T7 TAATACGACTCACTATAGG

```

```
Class: BAC ends
High quality sequence start: 16
High quality sequence stop: 847.
Location/Qualifiers
  1..968
    /organism="Xenopus tropicalis"
    /mol_type="genomic DNA"
    /db_xref="taxon:8364"
    /clone="ISB1-101G15"
    /clone_lib="ISB1"
    /note="Vector: pBelobAC11; ISB-1 Xenopus tropicalis BAC
    Library Segment 1"

FEATURES
    source
        1..968
            /organism="Xenopus tropicalis"
            /mol_type="genomic DNA"
            /db_xref="taxon:8364"
            /clone="ISB1-101G15"
            /clone_lib="ISB1"
            /note="Vector: pBelobAC11; ISB-1 Xenopus tropicalis BAC
            Library Segment 1"

ORIGIN
Query Match      67.9%; Score 852; DB 10; Length 968;
Best Local Similarity 99.2%; Pred. No. 4.5e-249;
Matches 877; Conservative 0; Mismatches 5; Indels 2; Gaps 2;

Qy 231 TCGAATTTCTGCCATTTCATCGCTTATATCACTTATTCAGCGGTAGCAACCGCGGTTT 290
Db 86 TCGAATTTCTGCCATTTCATCGCTTATATCACTTATTCAGCGGTAGCAACCGCGGTTT 145

Qy 291 AAGGCAACAATACTGCTTAAAAAATTAGCCCGCCCTGCCACTCATCGCACTACT 350
Db 146 AAGGCAACAATACTGCTTAAAAAATTAGCCCGCCCTGCCACTCATCGCACTACT 205

Qy 351 GTTGTAATTCATAAGCAATTCGCGACATGGAAGCCATCAAAACGGCATGTAACCT 410
Db 206 GTTGTAATTCATAAGCAATTCGCGACATGGAAGCCATCAAAACGGCATGTAACCT 265

Qy 411 GAATCGCAGCGGCATCGACCTTGTGCGCTTGCGTATATATTTGCCCATGGTGAAGA 470
Db 266 GAATCGCAGCGGCATCGACCTTGTGCGCTTGCGTATATATTTGCCCATGGTGAAGA 325

Qy 471 CGGGGGCGAAGAAGTTGTTCATATTTGGCCACCGCTTTAAATCAAACTGGTGAACCTCACCC 530
Db 326 CGGGGGCGAAGAAGTTGTTCATATTTGGCCACCGCTTTAAATCAAACTGGTGAACCTCACCC 385

Qy 531 AGGGATTTGGCTGAGACGAAAAACATATTTCTCAATAAACCTTTTAGGAAAAATAGGCCAGGT 590
Db 386 AGGGATTTGGCTGAGACGAAAAACATATTTCTCAATAAACCTTTTAGGAAAAATAGGCCAGGT 445

Qy 591 TTTCACCGTAACACGCCACATCTTCGGAATATATGTGTAGAACTGCCGGAATCGTCGT 650
Db 446 TTTCACCGTAACACGCCACATCTTCGGAATATATGTGTAGAACTGCCGGAATCGTCGT 505

Qy 651 GGTATTTCACTCCAGACGATGAAACGTTTCAAGTTTGCTTCATGGAACCGGTGTAACAAG 710
Db 506 GGTATTTCACTCCAGACGATGAAACGTTTCAAGTTTGCTTCATGGAACCGGTGTAACAAG 565

Qy 711 GGTGAACACTATCCCATATCACAGCTCACCGTCTTTTCATTTGCCCATACGTAATTCGGGA 770
Db 566 GGTGAACACTATCCCATATCACAGCTCACCGTCTTTTCATTTGCCCATACGTAATTCGGGA 625

Qy 771 GAGCATTTCATGCGGGGCAAGATGTGAATAAGCCGGATAAACTTGCTCTTATTTT 830
Db 626 GAGCATTTCATGCGGGGCAAGATGTGAATAAGCCGGATAAACTTGCTCTTATTTT 685

Qy 831 TCTTTACGCTCTTTAAAAAGCCGTAATATCCAGCTGGAACGCTCTGGTTATAGGTACATTT 890
Db 686 TCTTTACGCTCTTTAAAAAGCCGTAATATCCAGCTGGAACGCTCTGGTTATAGGTACATTT 745

Qy 891 GAGCAACTGACTGAAATGCTCAAAATGTTCTTTTAGATGCGCATTTGGGATATATCAACGG 950
Db 746 GAGCAACTGACTGAAATGCTCAAAATGTTCTTTTAGATGCGCATTTGGGATATATCAACGG 805

Qy 951 TGGTATATCCAGTGATTTTTTCTCCATTTAGCTTCTTAGCTCTCTGAAATCTCGACA 1010
Db 806 TGGTATATCCAGTGATTTTTTCTCCATTTAGCTTCTTAGCTCTCTGAAATCTCGACA 865

Qy 1011 ACTCAAAAAATACGCCCGGTAGTGATCTTATTTTCAATTATGGTGAAGTTT-GGAACCTCTTT 1069
Db 866 ACTCAAAAAATACGCCCGGTAGTGATCTTATTTTCAATTATGGTGAAGTTTGGGAACCTCTTT 925
```

```
Qy 1070 ACGTCCGATCAACGCTCTCATTTTTCGCAAAAGTTTGCCCGAGG 1113
Db 926 ACGTGGCGATC-ACGTCTCATTTTTCGCCAAAGTTGGCCCAAGG 968

RESULT 5
CL134875
LOCUS CL134875
DEFINITION ISB1-105K7_T7.1 ISB1 Xenopus tropicalis genomic clone ISB1-105K7,
genomic survey sequence.
ACCESSION CL134875
VERSION CL134875.1 GI:40628510
KEYWORDS GSS.
SOURCE Xenopus tropicalis (western clawed frog)
ORGANISM Xenopus tropicalis
Amphibia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Eukaryota; Mesozoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
Xenopodinae; Xenopus; Silurana.
REFERENCE 1 (bases 1 to 994)
AUTHORS Kremitzki,C., Carter,J., McPherson,J., Warren,W., Graves,T.,
Mardis,E. and Wilson,R.
TITLE A physical map of the xenopus tropicalis genome
JOURNAL Unpublished (2003)
COMMENT Contact: Richard K Wilson
Genome Sequencing Center
Washington University School of Medicine
Email: submissions@wustl.edu
Insert Length: 75000 Std Error: 0.00
Seq primer: T7 TAATACGACTCACTATAGG
Class: BAC ends
High quality sequence start: 6
High quality sequence stop: 804.
Location/Qualifiers
  1..994
    /organism="Xenopus tropicalis"
    /mol_type="genomic DNA"
    /db_xref="taxon:8364"
    /clone="ISB1-105K7"
    /clone_lib="ISB1"
    /note="Vector: pBelobAC11; ISB-1 Xenopus tropicalis BAC
    Library Segment 1"

ORIGIN
Query Match      67.9%; Score 852; DB 10; Length 994;
Best Local Similarity 97.5%; Pred. No. 4.5e-20;
Matches 887; Conservative 0; Mismatches 20; Indels 3; Gaps 2;

Qy 245 TTTCATCCGCTTATATCACTTATTCAGGCGTAGCAACCGAGGTTTAAGGCGCAACATAA 304
Db 1 TTCTGCCGCTTATATCACTTATTCAGGCGTAGCAACCGAGGTTTAAGGCGCAACATAA 60

Qy 305 CTGCTTAAAAAATTACGCCCGCCCTGCGACTCATCGAGTACTGTTGTAATTCATTA 364
Db 61 CTGCTTAAAAAATTACGCCCGCCCTGCGACTCATCGAGTACTGTTGTAATTCATTA 120

Qy 365 AGCATTTCGCGACATGGAAGCCATCAAAACGGCATGTAACCTGATGCCAGCGC 424
Db 121 AGCATTTCGCGACATGGAAGCCATCAAAACGGCATGTAACCTGATGCCAGCGC 180

Qy 425 ATCAGACCTTGTCCGCTTGGTATATAATTTGCCCATGGTGAACCGGCGAAGAAG 484
Db 181 ATCAGACCTTGTCCGCTTGGTATATAATTTGCCCATGGTGAACCGGCGAAGAAG 240

Qy 485 TTGTCATATTGGCCACGTTTAAATCAAACTGGTGAACCTCACCCAGGATTTGGCTGAG 544
Db 241 TTGTCATATTGGCCACGTTTAAATCAAACTGGTGAACCTCACCCAGGATTTGGCTGAG 300

Qy 545 ACGAAAAACATATTTCTCAATAAACCTTTAGGGAATAGCCAGGTTTTCACCGTAACAC 604
Db 301 ACGAAAAACATATTTCTCAATAAACCTTTAGGGAATAGCCAGGTTTTCACCGTAACAC 360

Qy 605 GCCACATCTTGGCAATATATGTAGAAACTGCCGGAATCGTCGTGGTATTTCACTCCAG 664
```

[illegible]

| | | | |
|------------|---|-----|--------|
| RESULT 6 | | | |
| CL132765 | | | |
| LOCUS | 1018 bp | DNA | linear |
| DEFINITION | CL132765. | | |
| | ISB1-10219_T7.1 ISB1 Xenopus tropicalis genomic clone ISB1-10219, | | |
| | genomic survey sequence. | | |

| ACCESSION | VERSION | KEYWORDS | SOURCE | ORGANISM |
|------------|---------|----------|---|----------|
| CU132765 | | | GI:40626400 | |
| CU132765.1 | | | GSS. | |
| | | | Xenopus tropicalis (western clawed frog) | |
| | | | Xenopus tropicalis | |
| | | | Xenopus tropicalis | |
| | | | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae; Xenopodinae; Xenopus; Silurana. | |

REFERENCE
AUTHORS
TITLE
JOURNAL
UNPUBLISHED (2003)

COMMENT

Contact: Richard K Wilson
 Genome Sequencing Center
 Washington University School of Medicine
 Email: submissions@watson.wustl.edu
 Insert length: 75000 Std Error: 0.00
 Ssq primer: T7 TAATACGACTCACTATAGGG

```

Class: BAC ends
High quality sequence start: 10
High quality sequence stop: 781.
Location/Qualifiers
source          1. .1018
FEATURES

```

```
/organism="Xenopus tropicalis"  
/mol_type="genomic DNA"  
/db_xref="taxon:8364"  
/clone="ISBL-10219"
```

10

```
RESULT 7
CL091480          1047 bp  DNA      linear      GSS 05-JAN-2004
LOCUS             ISB1-2019.T7.1 ISB1 Xenopus tropicalis genomic clone ISB1-2019,
DEFINITION        genomic survey sequence.
ACCESSION          CL091480
VERSION            CL091480.1  GI:40585115
KEYWORDS           GSS.
SOURCE             Xenopus tropicalis (western clawed frog)
ORGANISM           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
                   Xenopodinae; Xenopus; Silurana.
REFERENCE          1 (bases 1 to 1047)
AUTHORS            Kremitzki,C., Carter,J., McPherson,J., Warren,W., Graves,T.,
                   Mardis,E. and Wilson,R.
TITLE              A physical map of the xenopus tropicalis genome
JOURNAL            Unpublished (2003)
COMMENT            Contact: Richard K Wilson
                   Genome Sequencing Center
                   Washington University School of Medicine
                   Email: submissions@watson.wustl.edu
                   Insert Length: 75000 Std Error: 0.00
                   Seq primer: T7 TAATACGACTCACTATAGG
                   Claes: BAC ends
                   High quality sequence start: 29
                   High quality sequence stop: 769.
FEATURES           Location/Qualifiers
                     1..1047
                     /organism="Xenopus tropicalis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:8364"
                     /clone="ISB1-2019"
                     /clone_lib="ISB1"
                     /note="Vector: pBelobAC11; ISB-1 Xenopus tropicalis BAC
                     Library Segment 1"
ORIGIN
Query Match      66.8%; Score 838; DB 10; Length 1047;
Best Local Similarity 97.7%; Pred. No. 9e-245;
Matches 850; Conservative 0; Mismatches 20; Indels 0; Gaps 0;
QY 324 CCCGCCCTGCCACTCATCGAGTACTGTTGTAATTCATTAGCATTCTGCCGACATGA 383
DB 1 CCCGCCCTGCCACTCATCCCTTTACTGTTGTAATTCATTAGCATTCTGCCGACATGA 60
QY 384 AGCCATCAAAACGGCATCATGAACCTGAATCGCCAGCGGCATCAGCACCTTGTGCGCTT 443
DB 61 AGCCATCAAAACGGCATCATGAACCTGAATCGCCAGCGGCATCAGCACCTTGTGCGCTT 120
QY 444 GCGTATAATATTGGCCATGGTGAAGAACGGGGCGGAAGATTGTCCATATTGGCCACGT 503
DB 121 GCGTATAATATTGGCCATGGTGAAGAACGGGGCGGAAGATTGTCCATATTGGCCACGT 180
QY 504 TTAATCAAACTGTGTAACCTCAACCCAGGGATTGGCTGAGACGAAACATATTCTCA 563
DB 181 TTAATCAAACTGTGTAACCTCAACCCAGGGATTGGCTGAGACGAAACATATTCTCA 240
QY 564 TAAACCTTTAGGGAATAGGCGAGTTTTCACCGTAAACGCCACATCTTCGGAATATA 623
DB 241 TAAACCTTTAGGGAATAGGCGAGTTTTCACCGTAAACGCCACATCTTCGGAATATA 300
QY 624 TGTGTAGAACTGCCGGAATCGTCGTGGTATTTCATCTCAGAGCGATGAAAAAGTTTCAG 683
DB 301 TGTGTAGAACTGCCGGAATCGTCGTGGTATTTCATCTCAGAGCGATGAAAAAGTTTCAG 360
QY 684 TTTCATCGAAACCGGTGTAAACGGGTGAACACTATCCCATATCACCAGTCACT 743
DB 361 TTTCATCGAAACCGGTGTAAACGGGTGAACACTATCCCATATCACCAGTCACT 420
QY 744 CTTTCATGCCATACCTAAATTCGGATGAGCTTTCATCAGGCGGGCAAGAATGTGAATA 803
DB 421 CTTTCATGCCATACCTAAATTCGGATGAGCTTTCATCAGGCGGGCAAGAATGTGAATA 480
QY 804 AGGCCGATAAACCTTGTGCTTATTTTCTTAGGCTCTTTAAAGGCGCGTAATATCCA 863
DB 481 AGGCCGATAAACCTTGTGCTTATTTTCTTAGGCTCTTTAAAGGCGCGTAATATCCA 540
QY 864 GCTGAACGGCTGCTGTTATAGGTACATTGAGCAACTGACTGAAATGCTCAAAATGTTCTT 923
DB 541 GCTGAACGGCTGCTGTTATAGGTACATTGAGCAACTGACTGAAATGCTCAAAATGTTCTT 600
QY 924 TAGCATGCCATTGGGATATATCAACGGTGGTATATCCAGTGATTTTTTCTCCATTTTAG 983
DB 601 TAGCATGCCATTGGGATATATCAACGGTGGTATATCCAGTGATTTTTTCTCCATTTTAG 660
QY 984 CTTCTTAGCTCCCTGAAATCTCGACAACCTCAAAATACGCCGCTAGTATCTTATTT 1043
DB 661 CTTCTTAGCTCCCTGAAATCTCGACAACCTCAAAATACGCCGCTAGTATCTTATTT 720
QY 1044 CATTATGGTGAAGTTGGAACTCTTACGTGCCGATCAACGCTCTCATTTTCGCAAAAGT 1103
DB 721 CATTATGGTGAAGTTGGAACTCTTACGTGCCGATCAACGCTCTCATTTTCGCAAAAGT 780
QY 1104 TGGCCAGGGCTTCCGGTATCAACAGGACACAGAGATTTTATTTCTCGGAAGTAT 1163
DB 781 TGGCCAGGGCTTCCGGTATCAACAGGACACAGAGATTTTATTTATTTCTGCCAAATGAT 840
QY 1164 CTTCCGTACAGGTATTTATTCGGTTCGAA 1193
DB 841 CTTCCGTCCCGCAGGAATTTATTCGCCCAAA 870
RESULT 8
CL145346
LOCUS             ISB1-145L7.T7.1 ISB1 Xenopus tropicalis genomic clone ISB1-145L7,
DEFINITION        genomic survey sequence.
ACCESSION          CL145346
VERSION            CL145346.1  GI:40638981
KEYWORDS           GSS.
SOURCE             Xenopus tropicalis (western clawed frog)
ORGANISM           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
                   Xenopodinae; Xenopus; Silurana.
REFERENCE          1 (bases 1 to 1001)
AUTHORS            Kremitzki,C., Carter,J., McPherson,J., Warren,W., Graves,T.,
                   Mardis,E. and Wilson,R.
TITLE              A physical map of the xenopus tropicalis genome
JOURNAL            Unpublished (2003)
COMMENT            Contact: Richard K Wilson
                   Genome Sequencing Center
                   Washington University School of Medicine
                   Email: submissions@watson.wustl.edu
                   Insert Length: 75000 Std Error: 0.00
                   Seq primer: T7 TAATACGACTCACTATAGG
                   Claes: BAC ends
                   High quality sequence start: 10
                   High quality sequence stop: 776.
FEATURES           Location/Qualifiers
                     1..1001
                     /organism="Xenopus tropicalis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:8364"
                     /clone="ISB1-145L7"
                     /clone_lib="ISB1"
                     /note="Vector: pBelobAC11; ISB-1 Xenopus tropicalis BAC
                     Library Segment 1"
ORIGIN
Query Match      66.4%; Score 832.2; DB 10; Length 1001;
Best Local Similarity 96.9%; Pred. No. 5.3e-243;
Matches 890; Conservative 0; Mismatches 24; Indels 4; Gaps 4;
```

| | | | | |
|----|--------------------|-----------------|--------------------------------------|------|
| Qy | TCGAATTTCTGCCATTCA | TCGCCTTA | TTATCACTTTATTCAGGCGTAGCAACACGAGCGTTT | 290 |
| Db | TCGAATTTCTGCCATTCA | TCGCCTTA | TTATCACTTTATTCAGGCGTAGCAACACGAGCGTTT | 146 |
| Qy | AAGGGCACCAATAACTG | CCTTAAAAAATTA | AGCCCGCCCTGCCACTCATCGCAGTACT | 350 |
| Db | AAGGGCACCAATAACTG | CCTTAAAAAATTA | AGCCCGCCCTGCCACTCATCGCAGTACT | 206 |
| Qy | GTGTGAATTCATTAA | GCATTCGCGCATG | AGAAACCGCATGATGAACCT | 410 |
| Db | GTGTGAATTCATTAA | GCATTCGCGCATG | AGAAACCGCATGATGAACCT | 266 |
| Qy | GAATCGCGCAGGGCAT | CAGCACCTTTCG | CCCTTCGGTATAATATTTCGCCATGTTGAAAAA | 470 |
| Db | GAATCGCGCAGGGCAT | CATTACCTTTCG | CCCTTCGGTATAATATTTCGCCATGTTGAAAAA | 326 |
| Qy | CGGGGGCGAAGAGTTG | TGCCATATTGGC | CAGTTTAAATCAAACTGGTGAAACTCAACC | 533 |
| Db | CGGGGGCGAAGAGTTG | TGCCATATTGGC | CAGTTTAAATCAAACTGGTGAAACTCAACC | 386 |
| Qy | AGGATTTGGCTGAGA | CGAAAAACATATTC | CAATAAACCCTTTTAGGGAAAAATAGCCAGGT | 590 |
| Db | AGGATTTGGCTGAGA | CGAAAAACATATTC | CAATAAACCCTTTTAGGGAAAAATAGCCAGGT | 446 |
| Qy | TTTTACCGTAAACAG | CCACATCTTTCG | NAATATATGTAGAAAACTGCCGGAAAAATCTGTCGT | 650 |
| Db | TTTTACCGTAAACAG | CCACATCTTTCG | NAATATATGTAGAAAACTGCCGGAAAAATCTGTCGT | 506 |
| Qy | GGTATTCACCTCAG | AGCGATGAAACGTT | TCAGTTTGCCTCATGGAACCGGTGTAAACAAG | 710 |
| Db | GGTATTCACCTCAG | AGCGATGAAACGTT | TCAGTTTGCCTCATGGAACCGGTGTAAACAAG | 566 |
| Qy | GGTGAACACTATCC | CAATATCACC | AGCTCAACCTTTCATTCGCCATACGGAATTCGGAT | 770 |
| Db | GGTGAACACTATCC | CAATATCACC | AGCTCAACCTTTCATTCGCCATACGGAATTCGGAT | 626 |
| Qy | GAGCATTCATCAG | CGGGGCAAGATGT | GAATAAAGCGCGGATATAAACTTTGTGCTTATTTT | 830 |
| Db | GAGCATTCATCAG | CGGGGCAAGATGT | GAATAAAGCGCGGATATAAACTTTGTGCTTATTTT | 686 |
| Qy | TCCTTACGGTCTTTA | AAAAAGCCGTATAT | TCACGCTGAACCGGTCTGTTATAGGTACATT | 890 |
| Db | TCCTTACGGTCTTTA | AAAAAGCCGTATAT | TCACGCTGAACCGGTCTGTTATAGGTACATT | 745 |
| Qy | GAGCAACTGACTG | AAATGCCTCAAAATG | TCTTTACGATGCCATTTGGATATATCAACdG | 950 |
| Db | GAGCAACTGACTG | AAATGCCTCAAAATG | TCTTTACGATGCCATTTGGATATATCAACdG | 805 |
| Qy | TGTTATATCAGTGA | TTTTTTTTTCTC | CAATTTTAGCTTCCTGAAAAATCTCGACdA | 1010 |
| Db | TGTTATATCAGTGA | TTTTTTTTTCTC | CAATTTTAGCTTCCTGAAAAATCTCGACdA | 864 |
| Qy | ACTCAAAAAATACG | CCGGTAGTCACTTA | TTTCAATTATGTGAAAGTTGGAACTCTTtA | 1070 |
| Db | ACTCAAAAAATACG | CCGGTAGTCACTTA | TTTCAATTATGTGAAAGTTGGAACTCTTtA | 923 |
| Qy | CGTCCGATCAACGT | CTCATTTTCG | CAAAAAGTTGGCCAGGGCTTCCCGGTATCAACAG | 1130 |
| Db | CGTCCGATCAACGT | CTCATTTTCG | CAAAAAGTTGGCCAGGGCTTCCCGGTATCAACAG | 982 |
| Qy | GGACACAGAGTTAT | TTATTT | 1148 | |
| Db | GGACACAGAGTTAT | TTATTT | 1000 | |

| | | | | | | | |
|------------|------------|------------|------------|--------|-----|--------|-----------------|
| RESULT 9 | CL129946 | CL129946 | CL129946 | 971 bp | DNA | linear | GSS 05-JAN-2004 |
| LOCUS | CL129946 | CL129946 | CL129946 | 971 bp | DNA | linear | GSS 05-JAN-2004 |
| DEFINITION | CL129946 | CL129946 | CL129946 | 971 bp | DNA | linear | GSS 05-JAN-2004 |
| ACCESSION | CL129946 | CL129946 | CL129946 | 971 bp | DNA | linear | GSS 05-JAN-2004 |
| VERSION | CL129946.1 | CL129946.1 | CL129946.1 | 971 bp | DNA | linear | GSS 05-JAN-2004 |
| KEYWORDS | CL129946 | CL129946 | CL129946 | 971 bp | DNA | linear | GSS 05-JAN-2004 |

SOURCE Xenopus tropicalis (western clawed frog)
 ORGANISM Xenopus tropicalis
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
 Xenopodinae; Xenopus; Silurana.
 1 (bases 1 to 971)
 Kremenitzki, C., Carter, J., McPherson, J., Warren, W., Graves, T.,
 Mardis, E. and Wilson, R.
 A physical map of the xenopus tropicalis genome
 Unpublished (2003)
 CONTACT: Richard K Wilson
 Genome Sequencing Center
 Washington University School of Medicine
 Email: submissions@watson.wustl.edu
 Insert Length: 75000 Std Error: 0.00
 Seq primer: T7 TAATACGACTCACTATAGG
 Class: BAC ends
 High quality sequence start: 10
 High quality sequence stop: 786.
 FEATURES
 source
 1. .971
 /organism="Xenopus tropicalis"
 /mol_type="genomic DNA"
 /db_xref="taxon:8364"
 /clone_lib="ISB1-97H8"
 /clone_lib="ISB1"
 /note="Vector: pBelobAC11; ISB-1 Xenopus tropicalis BAC
 Library Segment 1"
 ORIGIN
 Query Match 66.3%; Score 831.4; DB 10; Length 971;
 Best Local Similarity 97.3%; Pred. No. 9.2e-243;
 Matches 877; Conservative 0; Mismatches 21; Indels 3; Gaps 3;
 QY 284 GCGCTTTAAGGCACCAATAA CTGCCCTTAAATAAATTCGCCCGCCCTGCCACTCATCG 343
 Db 3 GGTTTTTAAGGGCACCAATAA CTGCCCTTAAATAAATTCGCCCGCCCTGCCACTCTTCG 62
 QY 344 CAGTACTGTTGTAATTCATTAAAGCATCTGCCGACATGGAAGGCATCACAAACGGCATGA 403
 Db 63 TAGTACTGTTGTAATTCATTAAAGCATCTGCCGACATGGAAGGCATCACAAACGGCATGA 122
 QY 404 TGAACCTGAATGCCACGGGCATCAGCA CTTGTGCGCTTGGGTATAATAATTTGCCCATG 463
 Db 123 TGAACCTGAATGCCACGGGCATCAGCA CTTGTGCGCTTGGGTATAATAATTTGCCCATG 182
 QY 464 GTCAAAACGGGGCGGAAGAAGTGTCCATATTTGGCCACGTTTAAATCAAACTGGTGAAA 523
 Db 183 GTCAAAACGGGGCGGAAGAAGTGTCCATATTTGGCCACGTTTAAATCAAACTGGTGAAA 242
 QY 524 CTCACCCAGGATTTGGCTGAGACGAAACACATATTCCTCAATAAACCTTTAGGGAATAG 583
 Db 243 CTCACCCAGGATTTGGCTGAGACGAAACACATATTCCTTTTATTTTAGGGAATAG 302
 QY 584 GCCAGGTTTTTCCCGTAAACACGCCACATCTTGGCAATATATGTGTAGAAACTGCCGGAAA 643
 Db 303 GCCAGGTTTTTCCCGTAAACACGCCACATCTTGGCAATATATGTGTAGAAACTGCCGGAAA 362
 QY 644 TCGTCTGTGTATTCATCTCAGAGCGGATGAAGCGTTTCAGTTTGTCTCATGGAAAAACGGTG 703
 Db 363 TCGTCTGTGTATTCATCTCAGAGCGGATGAAGCGTTTCAGTTTGTCTCATGGAAAAACGGTG 422
 QY 704 TAAACAGGGTGAACACTATCCCATATCACAGCTCACCGCTCTTTTCAATTCGCCATACCGTAAT 763
 Db 423 TAAACAGGGTGAACACTATCCCATATCACAGCTCACCGCTCTTTTCAATTCGCCATACCGTAAT 482
 QY 764 TCCGGATGAGCATTTCA TCAGGCGGGCAAGAATGTGTAATAAGGCCGGATAAAACTTTGTGC 823
 Db 483 TCCGGATGAGCATTTCA TCAGGCGGGCAAGAATGTGTAATAAGGCCGGATAAAACTTTGTGC 542
 QY 824 TTATTTTCTTTTACGGTCTTTTAAAAAGGGCGGTAAATCCAGCTGAACGGTCTGTGGTTATAG 883
 Db 543 TTATTTTCTTTTACGGTCTTTTAAAAAGGGCGGTAAATCCAGCTGAACGGTCTGTGGTTATAG 602

```
Qy 884 GTACATTGAGCAACTGACTGAATGCTCBAATGCTTCTTTACGATGCCATTGGATATA 943
Db 603 GTACATTGAGCAACTGACTGAATGCTCBAATGCTTCTTTACGATGCCATTGGATATA 662
Qy 944 TCAACGGTGGTATATCCAGTGATTTTTTTCTCCCAATTTTAGCTTCTTAGCTCCTGAAAT 1003
Db 663 TCAACGGTGGTATATCCAGTGATTTTTTTCTCCCAATTTTAGCTTCTTAGCTCCTGAAAT 722
Qy 1004 CTCGACAACTCAAAAAATACCCCGGTAGTGATCTTATTTTCATTATGGTGAAGTTGAA 1063
Db 723 CTCGATACTCAAAAAATACCCCGGTAGTGATCTTATTTTCATTATGGTGAAGTTGAA 782
Qy 1064 CTTCTAGTCCGATCAAGCTCTCATTTTCGCCAAAGTTGGCCAGGC-TTCCCGGT 1122
Db 783 CTTCTAGTCCGATCAAGCTCTCATTTTCCCAAAGTTGGCCAGGC-TTCCCGGT 842
Qy 1123 ATCAACGGGACACCCAGGA-TTTATTTATCTGCGAAGTGAT-CTTCCGTACAGGTATT 1180
Db 843 ATCAACGGGACACCCAGGATTTATTTATTTCTGCGAAGTGATCTTCCGTCCAGGTATT 902
Qy 1181 T 1181
Db 903 T 903

RESULT 10
CL131806
LOCUS
DEFINITION
ISB1-101C13 T7.1 ISB1 Xenopus tropicalis genomic clone ISB1-101C13,
genomic survey sequence.
ACCESSION
CL131806
VERSION
CL131806.1 GI:40625441
KEYWORDS
GSS.
SOURCE
Xenopus tropicalis (western clawed frog)
ORGANISM
Xenopus tropicalis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
Xenopodinae; Xenopus; Silurana.
REFERENCE
1 (bases 1 to 968)
Kremitzki,C., Carter,J., McPherson,J., Warren,W., Graves,T.,
Mardis,E. and Wilson,R.
TITLE
A physical map of the xenopus tropicalis genome
JOURNAL
Unpublished (2003)
COMMENT
Contact: Richard K Wilson
Genome Sequencing Center
Washington University School of Medicine
Email: submissions@watson.wustl.edu
Insert Length: 75000 Std Error: 0.00
Seq primer: T7 TAATAGCTCACTATAGG
Class: BAC ends
High quality sequence start: 8
High quality sequence stop: 751.
FEATURES
Location/Qualifiers
source
1..968
/organism="Xenopus tropicalis"
/mol_type="genomic DNA"
/db_xref="taxon:8364"
/clone="ISB1-101C13"
/clone_lib="ISB1"
/vector="pBelBAC11; ISB-1 Xenopus tropicalis BAC
Library Segment 1"

ORIGIN
Query Match 66.2%; Score 829.8; DB 10; Length 968;
Best Local Similarity 97.7%; Pred. No. 2,9e-242;
Matches 862; Conservative 0; Mismatches 18; Indels 2; Gaps 2;
Qy 231 TCGAATTTCTGCATTCATCCCGCTTATATCACTTATTCAGCGGTAGCAACCGCGTTT 290
Db 86 TCGAATTTCTGCATTCATCCCGCTTATATCACTTATTCAGCGGTAGCAACCGCGTTT 145
Qy 291 AAGGGACCAATAACTGCTTAAAAAAATTTACGCCCGCCCTGCCACTCATCGCAGTACT 350
```

```
Db 146 AAGGGACCAATAACTGCTTAAAAAAATTTAGCCCCCGCTGCACTCATCGCAGTACT 205
Qy 351 GTTGTAAATTCAATTAAGCAATTCGCCGACATGGAAGCCATCACAAACGGCATGTAACCT 410
Db 206 GTTGTAAATTCAATTAAGCAATTCGCCGACATGGAAGCCATCACAAACGGCATGTAACCT 265
Qy 411 GNAATCGCCAGCGGCATCAGCACCTTGTGCGCTTGCCTGTATATAATTTTGCCTCATGTGAAA 470
Db 266 GAATCGCCAGCGGCATCAATTAACCTTGTGCGCTTGCCTGTATATAATTTTGCCTCATGTGAAA 325
Qy 471 CGGGGGGGAAGAAAGTTGTCCATATTTGCCACAGTTTAAATCAAAAATCTGCTGAAATCTACCC 530
Db 326 CGGGGGGGAAGAAAGTTGTCCATATTTGCCACAGTTTAAATCAAAAATCTGCTGAAATCTACCC 385
Qy 531 AGGGAATTTGGCTGAGACGAAACAAATATTTCTCAATAAACCTTTTAGGGAAATAGGCCAGGT 590
Db 386 AGGGAATTTGGCTGAGACGAAACAAATATTTCTCAATAAACCTTTTAGGGAAATAGGCCAGGT 445
Qy 591 TTTTCAACCGTAAACCGCCACATCTTTGCGAATATATGTGTAGAAACTGCGCGAAATCGTCTGT 650
Db 446 TTTTCAACCGTAAACCGCCACATCTTTGCGAATATATGTGTAGAAACTGCGCGAAATCGTCTGT 505
Qy 651 GGTATTCACTCCAGAGCGATGAAACCGTTTCAGTTTGTCTCATGGAACCGGTGTAAACAAG 710
Db 506 GGTATTCACTCCAGAGCGATGAAACCGTTTCAGTTTGTCTCATGGAACCGGTGTAAACAAG 565
Qy 711 GGTGAACACTATCCCATATCACAGCTCACCGTCTTTTCATTTGCCATAGTAAATTCGGAT 770
Db 566 GGTGAACACTATCCCATATCACAGCTCACCGTCTTTTCATTTGCCATAGTAAATTCGGAT 625
Qy 771 GAGCAATTCATCAGCGCGGCAAGAAATGTGAATAAAGCGGATAAAATCTTGCTTATTTT 830
Db 626 GAGCAATTCATCAGCGCGGCAAGAAATGTGAATAAAGCGGATAAAATCTTGCTTATTTT 685
Qy 831 TCTTTACGGTCTTTAAAAAGCCGCTTAATATCCAGCTGAACCGTCTGTTTATAGGTACATT 890
Db 686 TCTTTACGGTCTTTAAAAAGCCGCTTAATATCCAGCTGAACCGTCTGTTTATAGGTACATT 745
Qy 891 GAGCAATTCATGAAATGCTCAAAATGTTCTTTACGATGCCATGGATATATCAACGG 950
Db 746 GAGCAATTCATGAAATGCTCAAAATGTTCTTTACGATGCCATGGATATATCAACGG 805
Qy 951 TGGTATATCCAGTGAATTTTTTCTCCATTTTAGCTTCTTAGCTCCTGAAATCTCGACA 1010
Db 806 TGGTATATCCAGGATTTTTTTCTCCATTTTAGCTTCTTAGCTCCTGAAATCTCGATA 865
Qy 1011 ACTCAAAAAATACGCC-GGTAGTGATCTTATTTTCATTATCGTGAAGTTGGAACCTCTT 1069
Db 866 ACTCAAAAAATACGCCGGGAAGGATCTTATTTTCATTATGGGAAGTTGGACCTCTT 925
Qy 1070 ACGTG-CCGATCAAGCTCTCATTTTTCGCCAAAGTTGGCCCA 1110
Db 926 ACGTGCCGATCAAGCTCTCATTTTCTGCCAAAGTTGGGCCA 967

RESULT 11
CL112441
LOCUS
DEFINITION
ISB1-57C16 T7.1 ISB1 Xenopus tropicalis genomic clone ISB1-57C16,
genomic survey sequence.
ACCESSION
CL112441
VERSION
CL112441.1 GI:40606076
KEYWORDS
GSS.
SOURCE
Xenopus tropicalis (western clawed frog)
ORGANISM
Xenopus tropicalis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
Xenopodinae; Xenopus; Silurana.
REFERENCE
1 (bases 1 to 1007)
Kremitzki,C., Carter,J., McPherson,J., Warren,W., Graves,T.,
Mardis,E. and Wilson,R.
TITLE
A physical map of the xenopus tropicalis genome
```


| | |
|--------------------|--|
| JOURNAL COMMENT | Unpublished (2003) Contact: Richard K Wilson Genome Sequencing Center Washington University School of Medicine Email: submissions@watson.wustl.edu Insert Length: 75000 Std Error: 0.00 Seq primer: T7 TAATACGACTCATTAGGG Class: BAC ends High quality sequence start: 8 High quality sequence stop: 801. |
| FEATURES | Location/Qualifiers source 1..1007 /organism="Xenopus tropicalis" /mol_type="genomic DNA" /db_xref="taxon:8364" /clone="ISB1-57C16" /clone_lib="ISB1" /note="Vector: pBelOBAC11; ISB-1 Xenopus tropicalis BAC Library Segment 1" |
| ORIGIN | Query Match 65.8%; Score 824.6; DB 10; Length 1007; Best Local Similarity 95.4%; Pred. No. 1.1e-240; Matches 882; Conservative 0; Mismatches 39; Indels 4; Gaps 3; |
| Qy | 231 TCGAATTTCTGCCATTTCATCCCGTCTTATTATCACTTATTTCAGGCGTAGCAACCAGCGGT 290 |
| Db | 83 TCGAATTTCTGCCATTTCATCCCGTCTTATTATCACTTATTTCAGGCGTAGCAACCAGCGGT 142 |
| Qy | 291 AGGGCACCAATAAATGCCTTAATAAAATAGCCCGCCCTGCCCACATCATCGCAGTACT 350 |
| Db | 143 AGGGCACCAATAAATGCCTTAATAAAATAGCCCGCCCTGCCCACATCATCGCAGTACT 202 |
| Qy | 351 GTTGTAATTCATTAAAGCATTCGCGACATGAAGCCATCAACAACGGCATGTAACCT 410 |
| Db | 203 GTTGTAATTCATTAAAGCATTCGCGACATGAAGCCATCAACAACGGCATGTAACCT 262 |
| Qy | 411 GAATCCGCCAGCGCATCAGCACTTGTGCGCTTGGGTATAATATTGTCCTTGGTAAAA 470 |
| Db | 263 GAATCCGCCAGCGCATCAGCACTTGTGCGCTTGGGTATAATATTGTCCTTGGTAAAA 322 |
| Qy | 471 CGGGGCGAAGAAGTTGTCATATTGGCCACGTTTAAATCAAACCTGTGAAAATCACCC 530 |
| Db | 323 CGGGGCGAAGAAGTTGTCATATTGGCCACGTTTAAATCAAACCTGTGAAAATCACCC 382 |
| Qy | 531 AGGGATTGCTGAGACGAAACATATCTCAATAAACCTTTAGGGAATAGCCAGGT 590 |
| Db | 383 AGGGATTGCTGAGACGAAACATATCTCAATAAACCTTTAGGGAATAGCCAGGT 442 |
| Qy | 591 TTTCACCGTAACACGGCCATCTTGGCAATATATGTAGAAACTGCGGAAATGTCGT 650 |
| Db | 443 TTTCACCGTAACACGGCCATCTTGGCAATATATGTAGAAACTGCGGAAATGTCGT 502 |
| Qy | 651 GSTATTCACTCCAGACGATGAAAACGTTTCAGTTTGTCTCATGGAACCGGTGAACAAG 710 |
| Db | 503 GSTATTCACTCCAGACGATGAAAACGTTTTCAGTTTGTCTCATGGAACCGGTGAACAAG 562 |
| Qy | 711 GGTAACACATNTCCCNATATCAGCTACCGTCTTTCATTCGCAATAGTAATTCGGAT 770 |
| Db | 563 GGTAACACATNTCCCNATATCAGCTACCGTCTTTCATTCGCAATAGTAATTCGGAT 622 |
| Qy | 771 GAGCATTCATCAGGGGGCAAGATGTGAATAAAGCGCGATATAAATCTTGTCATTATT 830 |
| Db | 623 GAGCATTCATCAGGGGGCAAGATGTGAATAAAGCGCGATATAAATCTTGTCATTATT 682 |
| Qy | 831 TCTTTACGGTCTTTAAAAAGGCGGTAATATCCAGCTGAACCGTCTGGTTATAGGTACATT 890 |
| Db | 683 TCTTTACGGTCTTTAAAAAGGCGGTAATATCCAGCTGAACCGTCTGGTTATAGGTACATT 742 |
| Qy | 891 GAGCAACTGACTGAAATGCTCAAAATGTTCTTTTACATGTCATTTGGATATATCAACGG 950 |
| Db | 743 GAGCAACTGACTGAAATGCTCAAAATGTTCTTTTACATGTCATTTGGATATATCAACGG 802 |

| | |
|--------------------|---|
| JOURNAL COMMENT | Unpublished (2003) Contact: Richard K Wilson Genome Sequencing Center Washington University School of Medicine Email: submissions@watson.wustl.edu Insert Length: 75000 Std Error: 0.00 Seq primer: T7 TAATACGACTCATTAGGG Class: BAC ends Location/Qualifiers High quality sequence start: 8 High quality sequence stop: 801. |
| FEATURES | source 1..1007 /organism="Xenopus tropicalis" /mol_type="genomic DNA" /db_xref="taxon:8364" /clone="ISB1-57C16" /clone_lib="ISB1" /note="Vector: pBelOBAC11; ISB-1 Xenopus tropicalis BAC Library Segment 1" |
| ORIGIN | Query Match 65.8%; Score 824.6; DB 10; Length 1007; Best Local Similarity 95.4%; Pred. No. 1.1e-240; Matches 882; Conservative 0; Mismatches 39; Indels 4; Gaps 3; |
| Qy | 231 TCGAATTTCTGCCATTTCATCCCGTCTTATTATCACTTATTTCAGGCGTAGCAACCAGCGGT 290 |
| Db | 83 TCGAATTTCTGCCATTTCATCCCGTCTTATTATCACTTATTTCAGGCGTAGCAACCAGCGGT 142 |
| Qy | 291 AGGGCACCAATAAATGCCTTAATAAAATAGCCCGCCCTGCCCATCTATCGCAGTACT 350 |
| Db | 143 AGGGCACCAATAAATGCCTTAATAAAATAGCCCGCCCTGCCCATCTATCGCAGTACT 202 |
| Qy | 351 GTTGTAATTCATTAAAGCATTCGCGACATGAAGCCATCAACAACGGCATGTAACCT 410 |
| Db | 203 GTTGTAATTCATTAAAGCATTCGCGACATGAAGCCATCAACAACGGCATGTAACCT 262 |
| Qy | 411 GAATCCGCCAGCGCATCAGCACTTGTGCGCTTGGGTATAATATTGGCCCATGGTGAANA 470 |
| Db | 263 GAATCCGCCAGCGCATCAGCACTTGTGCGCTTGGGTATAATATTGGCCCATGGTGAANA 322 |
| Qy | 471 CGGGGCGGAAGAAGTTGTCATATTGGCCACGTTTAAATCAAAAATGTTGAAATCTCACCC 530 |
| Db | 323 CGGGGCGGAAGAAGTTGTCATATTGGCCACGTTTAAATCAAAAATGTTGAAATCTCACCC 382 |
| Qy | 531 AGGGATTGCTGAGACGAAACATATCTCAATAAACCTTTAGGGAATAGGCCAGGT 590 |
| Db | 383 AGGGATTGCTGAGACGAAACATATCTCAATAAACCTTTAGGGAATAGGCCAGGT 442 |
| Qy | 591 TTTCACCGTAACACGGCCATCTTGGCAATATATGTAGAAACTGCGGAAATGTCGT 650 |
| Db | 443 TTTCACCGTAACACGGCCATCTTGGCAATATATGTAGAAACTGCGGAAATGTCGT 502 |
| Qy | 651 GSTATTCACTCCAGACGCGATGAAACGTTTCAGTTTGTCTCATGGAANAACGGTGAACAAG 710 |
| Db | 503 GGTAATTCATCCAGACGCGATGAAACGTTTTCAGTTTGTCTCATGGAANAACGGTGAACAAG 562 |
| Qy | 711 GGTAACACATATCCCATATCACCAGTTCACGTCCTTTTCATTCGCGATAGTAATTCGGAT 770 |
| Db | 563 GGTAACACATATCCCATATCACCAGTTCACGTCCTTTTCATTCGCGATAGTAATTCGGAT 622 |
| Qy | 771 GAGCATTCATCAGGGGGCAAGATGTGAATAAAGCGCGATATAAATCTTGTCCTTATT 830 |
| Db | 623 GAGCATTCATCAGGGGGCAAGATGTGAATAAAGCGCGATATAAATCTTGTCCTTATT 682 |
| Qy | 831 TCTTTACGGTCTTTAAAAAGGCGGTAATATCCAGCTGAACGGTCTGGTTATAGGTACATT 890 |
| Db | 683 TCTTTACGGTCTTTAAAAAGGCGGTAATATCCAGCTGAACGGTCTGGTTATAGGTACATT 742 |
| Qy | 891 GAGCAACTGACTGAAATGCTCAAATGTTCTTTTACATGTCATTTGGGATATATCAACGG 950 |
| Db | 743 GAGCAACTGACTGAAATGCTCAAATGTTCTTTTACATGTCATTTGGGATATATCAACGG 802 |

| ORIGIN | | Library Segment 1" | | | | | | | | | |
|---------------------------|------|---|------|--|--|--|--|--|--|--|--|
| Query Match | | 64.1%; Score 803.6; DB 10; Length 901; | | | | | | | | | |
| Best Local Similarity | | 96.7%; Pred. No. 3e-234; | | | | | | | | | |
| Matches 853; Conservative | | 0; Mismatches 24; Indels 5; Gaps 3; | | | | | | | | | |
| Qy | 230 | TTGGAATTTTCGCCATTCATCGCTTATTATACACTTATTCAGGCGGTAGCAACCCAGGCGTT | 289 | | | | | | | | |
| Db | 3 | TCGGAAATTTTCGCCCTTCATCGCTTATTATACACTTATTCAGGCGGTAGCAACCCAGGCGTT | 62 | | | | | | | | |
| Qy | 290 | TAAGGGCACCAATAACTGCCTTTAAAAAAATTAACGCCCGCCCTGCCACTCATCGCAGTAC | 349 | | | | | | | | |
| Db | 63 | TCTGGCACCAATAACTGCCTTTAAAAAAATTAACGCCCGCCCTGCCACTCATCGCAGTAC | 122 | | | | | | | | |
| Qy | 350 | TGTTGTAATTCATTAAAGCAATTCGCGACATGGAAAGCCATCACAAACGGCATGATGAACC | 409 | | | | | | | | |
| Db | 123 | TGTTGTAATTCATTAAAGCAATTCGCGACATGGAAAGCCATCACAAACGGCATGATGAACC | 182 | | | | | | | | |
| Qy | 410 | TGAATCGCCAGCGCATCAGCACCTTGTCGCTTCGGTATATATTTGCCCATGTTGAAA | 469 | | | | | | | | |
| Db | 183 | TGAATCGCCAGCGCATCAGCACCTTGTCGCTTCGGTATATATTTGCCCATGTTGAAA | 242 | | | | | | | | |
| Qy | 470 | ACGGGGCGAAGAAGTTGTCCATATTGGCCACGTTTAAATCAAAACTGGTGAACCTCACC | 529 | | | | | | | | |
| Db | 243 | ACGGGGCGAAGAAGTTGTCCATATTGGCCACGTTTAAATCAAAACTGGTGAACCTCACC | 302 | | | | | | | | |
| Qy | 530 | CAGGATTTGGCTGAGACGAAACAAATATTCATATAAACCCCTTTAGGAAATAGGCCAGG | 589 | | | | | | | | |
| Db | 303 | CAGGATTTGGCTGAGACGAAACAAATATTCATATAAACCCCTTTAGGAAATAGGCCAGG | 362 | | | | | | | | |
| Qy | 590 | TTTTACCGTAACACGCCACATCTTGGCAATATATGTAGAACTGCCGGAATTCGTG | 649 | | | | | | | | |
| Db | 363 | TTTTACCGTAACACGCCACATCTTGGCAATATATGTAGAACTGCCGGAATTCGTG | 422 | | | | | | | | |
| Qy | 650 | TGATATTCACTCCAGAGCGATGAAAAAGTTTCAGTTGCTCATGGAAAAAGGTGTAAACA | 709 | | | | | | | | |
| Db | 423 | TGATATTCACTCCAGAGCGATGAAAAAGTTTCAGTTGCTCATGGAAAAAGGTGTAAACA | 482 | | | | | | | | |
| Qy | 710 | GGGTGAAACACTATCCCATATACCAAGCTCACCGTCTTTTCATTGCGCATACGTAATTCGG | 769 | | | | | | | | |
| Db | 483 | GGGTGAAACACTATCCCATATACCAAGCTCACCGTCTTTTCATTGCGCATACGTAATTCGG | 542 | | | | | | | | |
| Qy | 770 | TGAGCATTCATCAGCGGGCAAGATGTGAATAAGCGCGGATAAACTTGCTGCTTATTT | 829 | | | | | | | | |
| Db | 543 | TGAGCATTCATCAGCGGGCAAGATGTGAATAAGCGCGGATAAACTTGCTGCTTATTT | 602 | | | | | | | | |
| Qy | 830 | TTCTTTAGCGTCTTTAAAAAGCCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACAT | 889 | | | | | | | | |
| Db | 603 | TTCTTTAGCGTCTTTAAAAAGCCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACAT | 662 | | | | | | | | |
| Qy | 890 | TGAGCAACTGACTGAAATGCCCTCAAAATGTTCTTTTACGATGCCATTTGGGATATATCAACG | 949 | | | | | | | | |
| Db | 663 | TGAGCAACTGACTGAAATGCCCTCAAAATGTTCTTTTACGATGCCATTTGGGATATATCAACG | 722 | | | | | | | | |
| Qy | 950 | GTGGTATATCCAGTGAATTTTTCCTCAATTTTAGCTTCTTAGCTCCTGAAAACTCGAC | 1009 | | | | | | | | |
| Db | 723 | GTGGTATATCCAGTGAATTTTTCCTCAATTTTAGCTTCTTAGCTCCTGAAAACTCGAC | 782 | | | | | | | | |
| Qy | 1010 | AACTCAAAAAATACGCC-GGTAGTGATC-TTATTTCATTATGGTGAAGTTGGAACCTC | 1067 | | | | | | | | |
| Db | 783 | AACTCAAAAAATACGCCGGGTAGTGATCTTTTTTTCATTATGGTGAAGTTGGAACCTC | 842 | | | | | | | | |
| Qy | 1068 | TT---ACGTGCCGATCAACGTCATTTTCGCAAAAGTTGG | 1106 | | | | | | | | |
| Db | 843 | TTTACGTGCCGAATAAAGCTCTCTTTTTCCTCCCAAAAGTTG | 884 | | | | | | | | |

Search completed: January 18, 2006, 11:40:49
Job time : 5311.56 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2006 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 17, 2006, 18:06:53 ; Search time 257.623 Seconds
(without alignments)
11032.288 Million cell updates/sec

Title: US-10-511-327-5
Perfect score: 50
Sequence: 1 tgattgattgaacgaaac.....ttcacgataaatgcgaaac 50

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 5893141 seqs, 28421725653 residues 11766282
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl: *
1: gb.ba:*
2: gb.in:*
3: gb.env:*
4: gb.om:*
5: gb.ov:*
6: gb.pat:*
7: gb.ph:*
8: gb.pr:*
9: gb.ro:*
10: gb.sts:*
11: gb.sy:*
12: gb.un:*
13: gb.vi:*
14: gb.hg:*
15: gb.pl:*

ALIGNMENTS

RESULT 1
AR084420/c
LOCUS AR084420 Sequence 6 from patent US 5981177. 58 bp DNA linear PAT 01-SEP-2000
DEFINITION AR084420
ACCESSION AR084420
VERSION AR084420.1 GI:10011191
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 58)
AUTHORS Demirjian,D.C., Casadaban,M.J., Weber,J.Mark. and Gaines,G.L. III.
TITLE Protein fusion method and constructs
JOURNAL Patent: US 5981177-A 6 09-NOV-1999;
FEATURES Location/Qualifiers
source
1..58
/organism="unknown"
/mol_type="unassigned DNA"

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|--------------|--------------------|
| C 1 | 39.4 | 78.8 | 58 | 6 AR084420 | AR084420 Sequence |
| C 2 | 39.4 | 78.8 | 71 | 6 AR084419 | AR084419 Sequence |
| C 3 | 39.4 | 78.8 | 77 | 6 BD073246 | BD073246 In vitro |
| C 4 | 39.4 | 78.8 | 77 | 6 AR353866 | AR353866 Sequence |
| C 5 | 39.4 | 78.8 | 82 | 7 PMUBSR | M10177 Bacterioph |
| C 6 | 39.4 | 78.8 | 83 | 6 AR084418 | AR084418 Sequence |
| C 7 | 39.4 | 78.8 | 117 | 6 AR084417 | AR084417 Sequence |
| C 8 | 39.4 | 78.8 | 120 | 11 SYNGENE | M12560 Synthetic |
| C 9 | 39.4 | 78.8 | 150 | 7 PMUSEE1 | M10863 Bacterioph |
| C 10 | 39.4 | 78.8 | 220 | 7 NCWU3R | X05382 Mu-derived |
| C 11 | 39.4 | 78.8 | 220 | 7 PMUNE2 | M34920 Bacterioph |
| C 12 | 39.4 | 78.8 | 240 | 1 ECOTRPA | M33723 Escherichia |
| C 13 | 39.4 | 78.8 | 324 | 11 SYNPMUEND | M15949 Synthetic |
| C 14 | 39.4 | 78.8 | 903 | 6 A02708 | A02708 pMW506 DNA |
| C 15 | 39.4 | 78.8 | 1319 | 11 AY781404 | AY781404 Synthetic |
| C 16 | 39.4 | 78.8 | 1319 | 11 AY781404 | AY781404 Synthetic |
| C 17 | 39.4 | 78.8 | 1351 | 11 AY781405 | AY781405 Synthetic |
| C 18 | 39.4 | 78.8 | 1351 | 11 AY781405 | AY781405 Synthetic |

19 39.4 78.8 1524 11 AY781401
20 39.4 78.8 1524 11 AY781401
21 39.4 78.8 1808 11 AY781408
22 39.4 78.8 1808 11 AY781408
23 39.4 78.8 1937 11 AY781403
24 39.4 78.8 1937 11 AY781403
25 39.4 78.8 2037 11 AY781407
26 39.4 78.8 2037 11 AY781407
27 39.4 78.8 2062 11 AY781402
28 39.4 78.8 2062 11 AY781402
29 39.4 78.8 2162 11 AY781406
30 39.4 78.8 2162 11 AY781406
31 39.4 78.8 2413 11 AY781399
32 39.4 78.8 2413 11 AY781399
33 39.4 78.8 2933 11 AY781400
34 39.4 78.8 2933 11 AY781400
35 39.4 78.8 4666 1 ECOPOLBDA
36 39.4 78.8 4791 11 AY738638
37 39.4 78.8 4791 11 AY738638
38 39.4 78.8 7003 6 AR084426
39 39.4 78.8 15611 1 ECOPHNAQ
40 39.4 78.8 15611 6 AR229537
41 39.4 78.8 36717 7 AF083977
42 39.4 78.8 37199 11 AY860420
43 37.4 74.8 240 1 ECOTRPA
44 31.8 63.6 5755 1 ECOCYS
45 30.4 60.8 1925 1 ECOMALIXA

AY781401 Synthetic
AY781401 Synthetic
AY781408 Synthetic
AY781408 Synthetic
AY781403 Synthetic
AY781403 Synthetic
AY781407 Synthetic
AY781407 Synthetic
AY781402 Synthetic
AY781402 Synthetic
AY781406 Synthetic
AY781406 Synthetic
AY781399 Synthetic
AY781399 Synthetic
AY781400 Synthetic
AY781400 Synthetic
M35371 E.coli DNA
AY738638 Cloning v
AY738638 Cloning v
AR084426 Sequence
J05260 E.coli psid
AR229537 Sequence
AF083977 Bacterioph
AY860420 Cloning v
M33723 Escherichia
M32101 E.coli thio
M28539 E.coli mali

Query Match 78.8%; Score 39.4; DB 6; Length 58;
Best Local Similarity 97.6%; Pred. No. 0.0051;
Matches 40; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 10 GAACGAAAAACCGAAAGCGTTTCACGATAAATGCGAAAAAC 50
Db 49 GCACGAAAAACCGAAAGCGTTTCACGATAAATGCGAAAAAC 9
RESULT 2
AR084419/c
LOCUS AR084419 Sequence 5 from patent US 5981177. 71 bp DNA linear PAT 01-SEP-2000
DEFINITION AR084419
ACCESSION AR084419
VERSION AR084419.1 GI:10011190
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 71)
AUTHORS Demirjian,D.C., Casadaban,M.J., Weber,J.Mark. and Gaines,G.L. III.


```

FEATURES             Location/Qualifiers
     source           1..83
                     /organism="unknown"
                     /mol_type="unassigned DNA"

ORIGIN
Query Match          78.8%; Score 39.4; DB 6; Length 83;
Best Local Similarity 97.6%; Pred. No. 0.0049;
Matches 40; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 10 GAACGAAAAACGCGAAGCGTTTCACGATAAATGCGAAAC 50
    |
    |
    |
Db 74 GCACGAAAAACGCGAAGCGTTTCACGATAAATGCGAAAC 34
    |
    |
    |

RESULT 7
AR084417/c
LOCUS             AR084417             117 bp      DNA      linear      PAT 01-SEP-2000
DEFINITION       Sequence 3 from patent US 5981177.
ACCESSION        AR084417
VERSION          AR084417.1 GI:10011188
KEYWORDS
SOURCE           Unknown.
ORGANISM         Unclassified.
REFERENCE        1 (bases 1 to 117)
AUTHORS         Demirjian,D.C., Casadaban,M.J., Weber,J.Mark. and Gaines,G.L. III.
TITLE            Protein fusion method and constructs
JOURNAL          Patent: US 5981177-A 3 09-NOV-1999;
FEATURES         Location/Qualifiers
     source       1..117
                 /organism="unknown"
                 /mol_type="unassigned DNA"

ORIGIN
Query Match          78.8%; Score 39.4; DB 6; Length 117;
Best Local Similarity 97.6%; Pred. No. 0.0048;
Matches 40; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 10 GAACGAAAAACGCGAAGCGTTTCACGATAAATGCGAAAC 50
    |
    |
    |
Db 108 GCACGAAAAACGCGAAGCGTTTCACGATAAATGCGAAAC 68
    |
    |
    |

RESULT 8
SYNGENE
LOCUS             SYNGENE             120 bp      DNA      linear      SYN 27-APR-1993
DEFINITION       Synthetic gene utilizing phosphorylated synthetic fragments in vivo.
ACCESSION        M12560
VERSION          M12560.1 GI:208423
KEYWORDS         synthetic construct
SOURCE           synthetic construct
ORGANISM         other sequences; artificial sequences.
REFERENCE        1 (bases 1 to 120)
AUTHORS         Narang,S.A., Dubuc,G., Yao,F.L. and Michniewicz,J.J.
TITLE            'In vitro' method of assembling a synthetic gene
JOURNAL          Biochem. Biophys. Res. Commun. 134 (1), 407-411 (1986)
PUBMED          3004442
COMMENT          source text: Phophorylated synthetic DNA fragments.
FEATURES         Location/Qualifiers
     source       1..120
                 /organism="synthetic construct"
                 /mol_type="genomic DNA"
                 /db_xref="taxon:32630"

ORIGIN
Query Match          78.8%; Score 39.4; DB 11; Length 120;
Best Local Similarity 97.6%; Pred. No. 0.0048;
Matches 40; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 10 GAACGAAAAACGCGAAGCGTTTCACGATAAATGCGAAAC 50
    |
    |
    |

```

```

Db 15 GCACGAAAAACGCGAAGCGTTTCACGATAAATGCGAAAC 55
    |
    |
    |

RESULT 9
PMUSEE1
LOCUS             PMUSEE1             150 bp      DNA      linear      PHG 28-APR-1993
DEFINITION       Bacteriophage Mu DNA, SE end fragment.
ACCESSION        M10863
VERSION          M10863.1 GI:215593
KEYWORDS
SEGMENT          1 of 2
SOURCE           Enterobacteria phage Mu
ORGANISM         Enterobacteria phage Mu
REFERENCE        1 (bases 1 to 150)
AUTHORS         Allet,B.
TITLE            Nucleotide sequences at the ends of bacteriophage Mu DNA
JOURNAL          Nature 274 (5671), 553-558 (1978)
PUBMED          672985
COMMENT          Original source text: Bacteriophage Mu DNA.
FEATURES         Location/Qualifiers
     source       1..150
                 /organism="Enterobacteria phage Mu"
                 /mol_type="genomic DNA"
                 /db_xref="taxon:10677"

ORIGIN
Query Match          78.8%; Score 39.4; DB 7; Length 150;
Best Local Similarity 97.6%; Pred. No. 0.0048;
Matches 40; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 10 GAACGAAAAACGCGAAGCGTTTCACGATAAATGCGAAAC 50
    |
    |
    |
Db 18 GCACGAAAAACGCGAAGCGTTTCACGATAAATGCGAAAC 58
    |
    |
    |

RESULT 10
NCMU3R
LOCUS             NCMU3R             220 bp      DNA      linear      PHG 02-APR-1988
DEFINITION       Mu-derived phage lambda placMu3 provirus right end (MuR) inserted in pBRG1214.
ACCESSION        X05582
VERSION          X05582.1 GI:15446
KEYWORDS         Mu-like viruses.
SOURCE           Enterobacteria phage Mu
ORGANISM         Enterobacteria phage Mu
REFERENCE        1 (bases 1 to 220)
AUTHORS         Nag,D.K. and Berg,D.E.
TITLE            Specificity of bacteriophage Mu excision
JOURNAL          Mol. Gen. Genet. 207 (2-3), 395-401 (1987)
PUBMED          3039296
COMMENT          see X05580 for pBRG1214 with insertion sites
                 see X05581 for prophage left end.
FEATURES         Location/Qualifiers
     source       1..220
                 /organism="Enterobacteria phage Mu"
                 /mol_type="genomic DNA"
                 /db_xref="taxon:10677"
                 /note="right end"

ORIGIN
Query Match          78.8%; Score 39.4; DB 7; Length 220;
Best Local Similarity 97.6%; Pred. No. 0.0046;
Matches 40; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 10 GAACGAAAAACGCGAAGCGTTTCACGATAAATGCGAAAC 50
    |
    |
    |
Db 211 GCACGAAAAACGCGAAGCGTTTCACGATAAATGCGAAAC 171
    |
    |
    |

```

```
RESULT 11
PMUNE2/c
LOCUS          PMUNE2          220 bp      DNA          linear          PHG 28-APR-1993
DEFINITION     Bacteriophage mu genomic right end.
ACCESSION      M34920
VERSION        M34920.1 GI:215584
KEYWORDS       2 of 2
SEGMENT        Enterobacteria phage Mu
SOURCE         Enterobacteria phage Mu
ORGANISM       Viruses; dsDNA viruses, no RNA stage; Caudovirales; Myoviridae;
               Mu-like viruses
REFERENCE      1 (bases 1 to 220)
AUTHORS       Groenen,M.A., Timmers,E. and van de Putte,P.
TITLE         DNA sequences at the ends of the genome of bacteriophage Mu
               essential for transposition
JOURNAL       Proc. Natl. Acad. Sci. U.S.A. 82 (7), 2087-2091 (1985)
PUBMED        2984681
COMMENT       Original source text: Bacteriophage mu DNA.
FEATURES      Location/Qualifiers
               1..220
               /organism="Enterobacteria phage Mu"
               /mol_type="genomic DNA"
               /db_xref="taxon:10677"
ORIGIN
Query Match      78.8%; Score 39.4; DB 7; Length 220;
Best Local Similarity 97.6%; Pred. No. 0.0046;
Matches 40; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 10 GAACGAAAAACGCGAACGGTTTCACGATAAATCGGAAAAAC 50
| | | | | | | | | | | | | | | | | | | | | | | | | |
Db 211 GCACGAAAAACGCGAACGGTTTCACGATAAATCGGAAAAAC 171
| | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 12
ECOTRPA/c
LOCUS          ECOTRPA          240 bp      DNA          linear          BCT 21-JUN-2002
DEFINITION     Escherichia coli transposon Mu dl-R insertion site.
ACCESSION      M33723
VERSION        M33723.1 GI:208387
KEYWORDS       trp operon.
SOURCE         Escherichia coli
ORGANISM       Escherichia coli
               Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
               Enterobacteriaceae; Escherichia.
REFERENCE      1 (sites)
AUTHORS       Chang,C.N., Kuang,W.J. and Chen,E.Y.
TITLE         Nucleotide sequence of the alkaline phosphatase gene of Escherichia
               coli
JOURNAL       Gene 44 (1), 121-125 (1986)
PUBMED        3533724
REFERENCE      2 (bases 1 to 240)
AUTHORS       Metcalf,W., Steed,P.M. and Wanner,B.L.
TITLE         Identification of phosphate starvation-inducible genes in
               Escherichia coli K-12 by DNA sequence analysis of psi::lacZ(Mu dl)
               transcriptional fusions
JOURNAL       J. Bacteriol. 172 (6), 3191-3200 (1990)
PUBMED        2160940
FEATURES      Location/Qualifiers
               1..240
               /organism="Escherichia coli"
               /mol_type="genomic DNA"
               /db_xref="taxon:562"
               1..104
               /citation=[2]
               /transposon="Mu dl-R"
               <105..>240
               /note="trp operon"
               /codon_start=1
               /transl_table=11
repeat_region
CDS
/note="unnamed protein product; pNM506 mgl-B sequenz"
/codon_start=1
/transl_table=11

/protein_id="AAA75351.1"
/db_xref="GI:984720"
/translation="LRKGTWTLNPFYEGFGMYVPQILMPALRQLEAFVSAQKDP
E"
```

```
/protein_id="CAA00263.1"
/db_xref="GI:344653"
/translation="MNKKVLTLSAVMASLLFGAHAHAADTR"

ORIGIN
Query Match          78.8%; Score 39.4; DB 6; Length 903;
Best Local Similarity 97.6%; Pred. No. 0.0042;
Matches 40; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 10 GAACGAAAACGCGAAAGCGTTTCACGATAAATCGAAAC 50
    |||
Db 795 GCACGAAAACGCGAAAGCGTTTCACGATAAATCGAAAC 835

RESULT 15
AY781404
LOCUS          1319 bp DNA linear SYN 28-FEB-2005
DEFINITION Synthetic construct transposon mini-Mu transposon TnCR6, complete
sequence.
ACCESSION AY781404
VERSION AY781404.1 GI:60171997
KEYWORDS
SOURCE synthetic construct
ORGANISM synthetic construct
other sequences; artificial sequences.
REFERENCE 1 (bases 1 to 1319)
AUTHORS Zhang,C.
TITLE Direct Submission
JOURNAL Submitted (14-OCT-2004) Copyrat Pty Ltd, 27-31 Wright Street,
Clayton, Victoria 3168, Australia
FEATURES
    source
        1..1319
            /organism="synthetic construct"
            /mol_type="other DNA"
            /db_xref="taxon:32630"
    repeat_region
        1..1319
            /transposon="mini-Mu transposon TnCR6"
    CDS
        300..959
            /codon_start=1
            /transl_table=11
            /product="chloramphenicol acetyltransferase"
            /protein_id="AAI4471.1"
            /db_xref="GI:60171998"
            /translation="MEKKITGYTTVDISQWHRKEHFEAFQSAQCTYNQVLDITAF
            LKTVKNKHFPYAFIHILARLMAHPEFRMAKMDGLVWDSVHPCVTVFHEQETEF
            SSLWSEYHDDPQFLHIYSQDVACVGENLAYFPKGFIEKMPFVSANPWVSFTSLNV
            ANMDFEAPVETMKYTTQGDKVLMLPLAIQVHVAVCDFHVGRLNELQQYCDWQGG
            A"

ORIGIN
Query Match          78.8%; Score 39.4; DB 11; Length 1319;
Best Local Similarity 97.6%; Pred. No. 0.0041;
Matches 40; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 10 GAACGAAAACGCGAAAGCGTTTCACGATAAATCGAAAC 50
    |||
Db 15 GCACGAAAACGCGAAAGCGTTTCACGATAAATCGAAAC 55
```

Search completed: January 18, 2006, 04:25:17
Job time : 258.623 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.1.6
Copyright (c) 1993 - 2006 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 17, 2006, 17:02:23 ; Search time 31.092 Seconds
(without alignments)
10717.687 Million cell updates/sec

Title: US-10-511-327-5
Perfect score: 50
Sequence: 1 tgattgattgaacgaaac.....ttcagcataaatgcgaaac 50

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_21.*

- 1: Geneseqn1980s.*
- 2: Geneseqn1990s.*
- 3: Geneseqn2000s.*
- 4: Geneseqn2001as.*
- 5: Geneseqn2001bs.*
- 6: Geneseqn2002as.*
- 7: Geneseqn2002bs.*
- 8: Geneseqn2003as.*
- 9: Geneseqn2003bs.*
- 10: Geneseqn2003cs.*
- 11: Geneseqn2003ds.*
- 12: Geneseqn2004as.*
- 13: Geneseqn2004bs.*
- 14: Geneseqn2005s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|-------------|---------------------|
| 1 | 50 | 100.0 | 50 | 10 ACF58172 | AcF58172 Cat-Mu tr |
| 2 | 50 | 100.0 | 54 | 10 ACF58168 | AcF58168 Cat-Mu tr |
| 3 | 50 | 100.0 | 1254 | 10 ACF58169 | AcF58169 Modified |
| C 4 | 50 | 100.0 | 1254 | 10 ACF58169 | AcF58169 Modified |
| C 5 | 39.4 | 78.8 | 51 | 13 ADS13821 | AdS13821 Mu end DN |
| 6 | 39.4 | 78.8 | 54 | AD21279 | Ad21279 Precut tr |
| 7 | 39.4 | 78.8 | 54 | 10 ACF58171 | AcF58171 Cat-Mu (No |
| 8 | 39.4 | 78.8 | 54 | 10 ACF58170 | AcF58170 Cat-Mu tr |
| 9 | 39.4 | 78.8 | 56 | 13 ADS13820 | AdS13820 Mu end DN |
| C 10 | 39.4 | 78.8 | 58 | 2 AAZ28883 | Aaz28883 Mini-Mu t |
| C 11 | 39.4 | 78.8 | 71 | 2 AAZ28882 | Aaz28882 Mini-Mu t |
| 12 | 39.4 | 78.8 | 82 | 10 ACC80745 | Acc80745 Transposo |
| C 13 | 39.4 | 78.8 | 83 | 2 AAZ28881 | Aaz28881 Mini-Mu t |
| 14 | 39.4 | 78.8 | 83 | 10 ACC80741 | Acc80741 Transposo |
| 15 | 39.4 | 78.8 | 84 | 10 ACC80738 | Acc80738 Transposo |
| 16 | 39.4 | 78.8 | 86 | 10 ACC80742 | Acc80742 Transposo |
| 17 | 39.4 | 78.8 | 89 | 10 ACC80753 | Acc80753 Transposo |
| 18 | 39.4 | 78.8 | 90 | 13 ADRI4827 | AdRI4827 PCR prime |
| C 19 | 39.4 | 78.8 | 117 | 2 AAZ28880 | Aaz28880 Mini-Mu t |

| | | | | | | |
|------|------|------|--------|----|----------|--------------------|
| 20 | 39.4 | 78.8 | 903 | 1 | AA80102 | Aa80102 EcoRI/Bam |
| C 21 | 39.4 | 78.8 | 7003 | 2 | AA228877 | Aa228877 Transposo |
| 22 | 39.4 | 78.8 | 15611 | 3 | AA01008 | Aa01008 Escherich |
| 23 | 28 | 56.0 | 30 | 4 | AA21278 | Aa21278 Bacteriop |
| C 24 | 25 | 50.0 | 34 | 14 | AE45577 | Aeb45577 SigA2 wit |
| C 25 | 25 | 50.0 | 34 | 14 | AE48763 | Aeb48763 Beta-lact |
| C 26 | 24.6 | 49.2 | 1874 | 6 | AA98204 | Aa98204 DNA encod |
| C 27 | 24.6 | 49.2 | 2229 | 7 | AD573196 | Ad573196 Human kid |
| C 28 | 24.6 | 49.2 | 2229 | 7 | ADW42050 | Adw42050 cDNA elev |
| C 29 | 24.6 | 49.2 | 6730 | 11 | ADL22564 | Adl22564 Human dis |
| C 30 | 24.6 | 49.2 | 17687 | 4 | AAK71665 | Aak71665 Human imm |
| C 31 | 24.6 | 49.2 | 17687 | 4 | AAK64966 | Aak64966 Human imm |
| C 32 | 24.6 | 49.2 | 17979 | 4 | AAK71664 | Aak71664 Human imm |
| C 33 | 24.6 | 49.2 | 17979 | 4 | AAK64964 | Aak64964 Human imm |
| C 34 | 24 | 48.0 | 2000 | 11 | ACL38625 | AcL38625 Rice stre |
| C 35 | 24 | 48.0 | 2099 | 6 | ABQ69021 | AbQ69021 Listeria |
| C 36 | 24 | 48.0 | 7868 | 6 | ABQ71053 | AbQ71053 Listeria |
| C 37 | 24 | 48.0 | 110000 | 6 | ABQ69245 | AbQ69245 09 |
| C 38 | 24 | 48.0 | 110000 | 6 | ABQ67197 | AbQ67197 08 |
| 39 | 23.8 | 47.6 | 2022 | 13 | ADR85671 | AdR85671 Aspergill |
| 40 | 23.8 | 47.6 | 2022 | 13 | ADR85084 | AdR85084 Aspergill |
| 41 | 23.8 | 47.6 | 8022 | 13 | ADR84497 | AdR84497 Aspergill |
| 42 | 23.6 | 47.2 | 917 | 6 | ABQ25453 | AbQ25453 Oligonuel |
| C 43 | 23.6 | 47.2 | 917 | 6 | ABQ25452 | AbQ25452 Oligonuel |
| C 44 | 23.6 | 47.2 | 955 | 6 | ABQ32590 | AbQ32590 Oligonuel |
| 45 | 23.6 | 47.2 | 955 | 6 | ABQ32591 | AbQ32591 Oligonuel |

ALIGNMENTS

RESULT 1

ACF58172
ID ACF58172 standard; DNA; 50 BP.

XX ACF58172;

DT 15-JAN-2004 (first entry)

DE Cat-Mu transposon modified end fragment.

XX Transposon; genetic engineering; transposase; Cat-Mu; ds.

OS Synthetic.

OS Bacteriophage mu.

PN WO2003087370-A1.

PD 23-OCT-2003.

PF 14-APR-2003; 2003WO-FI000285.

PR 18-APR-2002; 2002FI-00000746.

XX (FINN-) FINNZYMES OX.

XX Savilahti H, Tiesaho V;

XX WPI; 2003-845329/78.

XX New transposon nucleic acid comprising a genetically engineered translation stop signal within a transposon end sequence recognized by a transposase useful for producing deletion derivatives of polypeptide.

XX Claim 6; Page 31; Opp; English.

XX The invention relates to a transposon nucleic acid comprising a genetically engineered translation stop signal in the three reading frames at least partly within a transposon end sequence recognized by a transposase. The transposon is useful for producing deletion derivatives of polypeptide coding nucleic acids. The method involves performing a transposition reaction in the presence of a target nucleic acid containing a polypeptide coding nucleic acid and in the presence of a

CC transposon containing a genetically engineered translation stop signal
 CC sequence in three reading frames at least partly within a transposon end
 CC sequence recognized by a transposase; and recovering a target nucleic
 CC acid having the transposon incorporated in the protein coding nucleic
 CC acid. The present sequence represents a Cat-Mu transposon modified end
 CC fragment without 5' overhang
 XX
 SQ Sequence 50 BP; 21 A; 8 C; 11 G; 10 T; 0 U; 0 Other;
 Query Match 100.0%; Score 50; DB 10; Length 50;
 Best Local Similarity 100.0%; Pred. No. 9.5e-09;
 Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 TGATTGATTGAACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAAC 50
 Db 1 TGATTGATTGAACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAAC 50
 RESULT 2
 ACF58168
 ID ACF58168 standard; DNA; 54 BP.
 AC ACF58168;
 XX
 XX
 DT 15-JAN-2004 (first entry)
 XX
 DE Cat-Mu transposon modified end fragment.
 XX
 KW Transposon; genetic engineering; transposase; Cat-Mu; ds.
 XX
 OS Synthetic.
 OS Bacteriophage mu.
 XX
 FH Key Location/Qualifiers
 FT misc_feature 1..4
 FT /*tag= a
 FT /note= "the 5' end of this strand overhangs the 3' end of
 FT the complementary strand"
 XX
 PN WO2003087370-A1.
 XX
 PD 23-OCT-2003.
 XX
 PF 14-APR-2003; 2003WO-FI000285.
 XX
 PR 18-APR-2002; 2002FI-00000746.
 XX
 PA (FINN-) FINNZYMES OY.
 XX
 PI Savilahti H, Tieaho V;
 XX WPI; 2003-845329/78.
 XX
 PT New transposon nucleic acid comprising a genetically engineered
 PT translation stop signal within a transposon end sequence recognized by a
 PT transposase useful for producing deletion derivatives of polypeptide.
 XX
 PS Claim 6; Fig 2; Opp; English.
 XX
 CC The invention relates to a transposon nucleic acid comprising a
 CC genetically engineered translation stop signal in the three reading
 CC frames at least partly within a transposon end sequence recognized by a
 CC transposase. The transposon is useful for producing deletion derivatives
 CC of polypeptide coding nucleic acids. The method involves performing a
 CC transposition reaction in the presence of a target nucleic acid
 CC containing a polypeptide coding nucleic acid and in the presence of a
 CC transposon containing a genetically engineered translation stop signal
 CC sequence in three reading frames at least partly within a transposon end
 CC acid having the transposon incorporated in the protein coding nucleic
 CC acid. The present sequence represents a Cat-Mu transposon modified end
 CC fragment
 XX

SQ Sequence 54 BP; 22 A; 9 C; 12 G; 11 T; 0 U; 0 Other;
 Query Match 100.0%; Score 50; DB 10; Length 54;
 Best Local Similarity 100.0%; Pred. No. 9.5e-09;
 Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 TGATTGATTGAACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAAC 50
 Db 5 TGATTGATTGAACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAAC 54
 RESULT 3
 ACF58169
 ID ACF58169 standard; DNA; 1254 BP.
 XX
 AC ACF58169;
 XX
 DT 15-JAN-2004 (first entry)
 XX
 DE Modified Cat-Mu(Stop)-transposon.
 XX
 KW Transposon; genetic engineering; transposase; Cat-Mu; ds.
 XX
 OS Synthetic.
 OS Bacteriophage mu.
 XX
 PN WO2003087370-A1.
 XX
 PD 23-OCT-2003.
 XX
 PF 14-APR-2003; 2003WO-FI000285.
 XX
 PR 18-APR-2002; 2002FI-00000746.
 XX
 PA (FINN-) FINNZYMES OY.
 XX
 PI Savilahti H, Tieaho V;
 XX WPI; 2003-845329/78.
 XX
 PT New transposon nucleic acid comprising a genetically engineered
 PT translation stop signal within a transposon end sequence recognized by a
 PT transposase useful for producing deletion derivatives of polypeptide.
 XX
 PS Claim 6; Page 30; Opp; English.
 XX
 CC The invention relates to a transposon nucleic acid comprising a
 CC genetically engineered translation stop signal in the three reading
 CC frames at least partly within a transposon end sequence recognized by a
 CC transposase. The transposon is useful for producing deletion derivatives
 CC of polypeptide coding nucleic acids. The method involves performing a
 CC transposition reaction in the presence of a target nucleic acid
 CC containing a polypeptide coding nucleic acid and in the presence of a
 CC transposon containing a genetically engineered translation stop signal
 CC sequence in three reading frames at least partly within a transposon end
 CC acid having the transposon incorporated in the protein coding nucleic
 CC acid. The present sequence represents a modified Cat-Mu(Stop)-transposon
 XX
 SQ Sequence 1254 BP; 356 A; 296 C; 264 G; 338 T; 0 U; 0 Other;
 Query Match 100.0%; Score 50; DB 10; Length 1254;
 Best Local Similarity 100.0%; Pred. No. 1.4e-08;
 Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 TGATTGATTGAACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAAC 50
 Db 5 TGATTGATTGAACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAAC 54
 RESULT 4
 ACF58169/c
 ID ACF58169 standard; DNA; 1254 BP.

```

XX ACF58169;
AC
XX
DT 15-JAN-2004 (first entry)
XX
DE Modified Cat-Mu(Stop)-transposon.
XX
KW Transposon; genetic engineering; transposase; Cat-Mu; ds.
XX
XX Synthetic.
OS Bacteriophage mu.
XX WO2003087370-A1.
XX
PD 23-OCT-2003.
XX
XX 14-APR-2003; 2003WO-FI000285.
XX
XX 18-APR-2002; 2002FI-00000746.
XX
XX (FINN-) FINNZYMES OY.
XX
XX Savilahti H, Tieaho V;
XX WPI; 2003-845329/78.
XX
XX New transposon nucleic acid comprising a genetically engineered
PT translation stop signal within a transposon end sequence recognized by a
PT transposase useful for producing deletion derivatives of polypeptide.
XX
XX Claim 6; Page 30; Opp; English.
XX
XX The invention relates to a transposon nucleic acid comprising a
CC genetically engineered translation stop signal in the three reading
CC frames at least partly within a transposon end sequence recognized by a
CC transposase. The transposon is useful for producing deletion derivatives
CC of polypeptide coding nucleic acids. The method involves performing a
CC transposition reaction in the presence of a target nucleic acid
CC containing a polypeptide coding nucleic acid and in the presence of a
CC transposon containing a genetically engineered translation stop signal
CC sequence in three reading frames at least partly within a transposon end
CC sequence recognized by a transposase; and recovering a target nucleic
CC acid having the transposon incorporated in the protein coding nucleic
CC acid. The present sequence represents a modified Cat-Mu(Stop)-transposon
XX
XX Sequence 1254 BP; 356 A; 296 C; 264 G; 338 T; 0 U; 0 Other;
SQ
Query Match 100.0%; Score 50; DB 10; Length 1254;
Best Local Similarity 100.0%; Pred. No. 1.4e-08;
Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 TGATTGATTGAACGAAACGCGAAGCGTTTCACGATAAATGCGAAAC 50
DB 1250 TGATTGATTGAACGAAACGCGAAGCGTTTCACGATAAATGCGAAAC 1201
RESULT 5
ADSI3821/c
ID ADSI3821 standard; DNA; 51 BP.
XX
AC ADSI3821;
XX
XX 16-DEC-2004 (first entry)
DT
XX
DE Mu end DNA fragment construction oligonucleotide MM1141.
XX
XX ss; mismatch detection; transposition detection; pathogen typing;
KW embryo screening; mutation detection; Mu end DNA; MM1141.
XX
XX Enterobacteria phage Mu.
OS Synthetic.
XX
XX US2004191821-A1.
FN
XX
XX 30-SEP-2004.
PD
XX
XX 26-MAR-2004; 2004US-00809688.
PF
XX
XX 28-MAR-2003; 2003US-0457934P.
PR
XX
XX (USSA ) US SEC OF ARMY.
PA
XX Yanagihara K, Mizuuchi K;
XX WPI; 2004-689846/67.
XX
XX Detecting a mismatch in a test double stranded nucleic acid target,
PT useful for typing a pathogenic microorganism strain, comprises detecting
PT transposition of the Mu-end nucleic acid into the target.
XX
XX Example 1; SEQ ID NO 2; 24pp; English.
XX
XX The invention relates to a method of detecting a mismatch in a test
CC double stranded nucleic acid target which comprises detecting
CC transposition of the Mu-end nucleic acid into the target, where
CC transposition at the predominant site indicates the presence of mismatch
CC at about that site. The methods are useful for typing a pathogenic
CC microorganism strain, for screening an embryo for the presence of
CC mutation, for detecting the presence of known mutation in a gene of
CC interest, detecting the presence of a previously unidentified mutation in
CC a gene of interest, and diagnosing the presence or absence of a tumour-
CC promoting mutation. The kit is useful for detecting the presence of a
CC mutation or polymorphism of interest in a nucleic acid molecule. The
CC present sequence represents the Mu end DNA fragment construction
CC oligonucleotide MM1141.
XX
XX Sequence 51 BP; 6 A; 13 C; 11 G; 21 T; 0 U; 0 Other;
SQ
Query Match 78.8%; Score 39.4; DB 13; Length 51;
Best Local Similarity 97.6%; Pred. No. 7.5e-05;
Matches 40; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 10 GAACGAAACGCGAAGCGTTTCACGATAAATGCGAAAC 50
DB 42 GCACGAAACGCGAAGCGTTTCACGATAAATGCGAAAC 2
RESULT 6
AAD21279
ID AAD21279 standard; DNA; 54 BP.
XX
AC AAD21279;
XX
XX 11-SEP-2003 (revised)
DT
XX 28-JAN-2002 (first entry)
DT
XX
DE Prescut transposon end of Bacteriophage Mu non-transferred strand.
XX
XX Insertional mutation; synaptic complex; transposon; screening; ds.
KW
XX Enterobacteria phage Mu.
OS
XX US6294385-B1.
XX
XX 25-SEP-2001.
PD
XX
XX 10-AUG-2000; 2000US-00615969.
PF
XX
XX 23-SEP-1998; 98US-00159363.
PR
XX
XX (WISC ) WISCONSIN ALUMNI RES FOUND.
PA
XX Goryshin IY, Reznikoff WS;
XX WPI; 2001-656171/75.
XX

```


CC acid having the transposon incorporated in the protein coding nucleic
 CC acid. The present sequence represents a Cat-Mu transposon containing wild
 CC -type Mu ends
 XX Sequence 54 BP; 21 A; 12 C; 14 G; 7 T; 0 U; 0 Other;
 SQ Query Match 78.8%; Score 39.4; DB 10; Length 54;
 Best Local Similarity 97.6%; Pred. No. 7.5e-05;
 Matches 40; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 10 GAACGAAACCGGAAAGCGTTTCACGATAAATGCGAAAC 50
 DB 14 GCACGAAACCGGAAAGCGTTTCACGATAAATGCGAAAC 54
 RESULT 9
 ADS13820
 ID ADS13820 standard; DNA; 56 BP.
 XX
 AC ADS13820;
 XX
 DT 16-DEC-2004 (first entry)
 XX
 DE Mu end DNA fragment construction oligonucleotide MM1138.
 XX
 KW ss; mismatch detection; transposition detection; pathogen typing;
 KW embryo screening; mutation detection; Mu end DNA; MM1138.
 XX
 OS Enterobacteria phage Mu.
 OS Synthetic.
 XX
 PN US2004191821-A1.
 XX
 PD 30-SEP-2004.
 XX
 PF 26-MAR-2004; 2004US-00809688.
 XX
 PR 28-MAR-2003; 2003US-0457934P.
 XX
 PA (USSA) US SEC OF ARMY.
 XX
 PI Yanagihara K, Mizuuchi K;
 XX
 DR WPI; 2004-689846/67.
 XX
 CC Detecting a mismatch in a test double stranded nucleic acid target,
 CC useful for typing a pathogenic microorganism strain, comprises detecting
 CC transposition of the Mu-end nucleic acid into the target.
 XX
 PS Example 1; SEQ ID NO 1; 24pp; English.
 XX
 CC The invention relates to a method of detecting a mismatch in a test
 CC double stranded nucleic acid target which comprises detecting
 CC transposition of the Mu-end nucleic acid into the target, where
 CC transposition at the predominant site indicates the presence of mismatch
 CC at about that site. The methods are useful for typing a pathogenic
 CC microorganism strain, for screening an embryo for the presence of
 CC mutation, for detecting the presence of known mutation in a gene of
 CC interest, detecting the presence of a previously unidentified mutation in
 CC a gene of interest, and diagnosing the presence or absence of a tumour-
 CC promoting mutation. The kit is useful for detecting the presence of a
 CC mutation or polymorphism of interest in a nucleic acid molecule. The
 CC present sequence represents the Mu end DNA fragment construction
 CC oligonucleotide MM1138.
 XX
 SQ Sequence 56 BP; 22 A; 12 C; 15 G; 7 T; 0 U; 0 Other;
 Query Match 78.8%; Score 39.4; DB 13; Length 56;
 Best Local Similarity 97.6%; Pred. No. 7.6e-05;
 Matches 40; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 10 GAACGAAACCGGAAAGCGTTTCACGATAAATGCGAAAC 50
 DB 14 GCACGAAACCGGAAAGCGTTTCACGATAAATGCGAAAC 54

DB 15 GCACGAAACCGGAAAGCGTTTCACGATAAATGCGAAAC 55
 RESULT 10
 AAZ28883/c
 ID AAZ28883 standard; cDNA; 58 BP.
 XX
 AC AAZ28883;
 XX
 DT 15-FEB-2000 (first entry)
 XX
 DE Mini-Mu transposable element deletion region delta-1.
 XX
 KW Transposon Mu; transposable element; fusion protein; attachment site;
 KW attL; attR; protein domain library; enzyme; accelerated evolution; ss.
 XX
 OS Synthetic.
 XX
 PN US5981177-A.
 XX
 PD 09-NOV-1999.
 XX
 PF 25-JAN-1995; 95US-00378548.
 XX
 PR 25-JAN-1995; 95US-00378548.
 XX
 PA (DEMI/) DEMIRJIAN D C.
 PA (CASA/) CASADABAN M J.
 PA (WEBE/) WEBER J M.
 PA (GAIN/) GAINES G L.
 XX
 PI Casadaban MJ, Demirjian DC, Weber JM, Gaines GL;
 XX
 DR WPI; 1999-633307/54.
 XX
 PT Generating fusion proteins using transposable elements, useful for
 PT development of a protein domain library and in the construction of multi-
 PT functional enzymes.
 XX
 PS Example 2; Fig 2; 41pp; English.
 XX
 CC The invention relates to a Mu-like transposable element (I) used for
 CC generating functional fusion proteins after insertion into a target DNA.
 CC The Mu-like element comprises: (a) a left transposable element attachment
 CC site attL and a right transposable element attachment site attR, where
 CC attR is no more than 62 nucleotides long; (b) a site for insertion of an
 CC exogenous DNA sequence encoding for a protein domain located between attL
 CC and attR; (c) after insertion of the transposable element into a target
 CC DNA sequence, a fusion mRNA sequence is transcribed originating either
 CC from the target DNA on either side of the transposable element or from
 CC inside the transposable element and continuing through the attachment
 CC site sequences and into the protein coding region, resulting in a single
 CC fusion open reading frame (ORF). The constructs are useful in the
 CC development of a protein domain library, in the construction of multi-
 CC functional enzymes and in the accelerated evolution of new enzymatic
 CC activities. The sequences AAZ28880-228884 represent deletion mini-Mu
 CC elements of the invention (encoded ORF - AAY55901-Y55906)
 XX
 SQ Sequence 58 BP; 8 A; 15 C; 13 G; 22 T; 0 U; 0 Other;
 Query Match 78.8%; Score 39.4; DB 2; Length 58;
 Best Local Similarity 97.6%; Pred. No. 7.6e-05;
 Matches 40; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 10 GAACGAAACCGGAAAGCGTTTCACGATAAATGCGAAAC 50
 DB 49 GCACGAAACCGGAAAGCGTTTCACGATAAATGCGAAAC 9
 RESULT 11
 AAZ28882/c
 ID AAZ28882 standard; cDNA; 71 BP.
 XX

AC AAZ28882;
 XX
 DT 15-FEB-2000 (first entry)
 XX
 XX Mini-Mu transposable element deletion region delta-21.
 XX
 KW Transposon Mu; transposable element; fusion protein; attachment site;
 KW attL; attR; protein domain library; enzyme; accelerated evolution; ss.
 XX
 OS Synthetic.
 XX
 XX US5981177-A.
 XX
 PN 09-NOV-1999.
 PD
 XX
 XX 25-JAN-1995; 95US-00378548.
 XX
 XX 25-JAN-1995; 95US-00378548.
 XX
 XX (DEMI/) DEMIRJIAN D C.
 PA (CASA/) CASADABAN M J.
 PA (WEBE/) WEBER J M.
 PA (GAIN/) GAINES G L.
 XX
 XX Casadaban MJ, Demirjian DC, Weber JM, Gaines GL;
 PI
 XX WPI; 1999-633307/54.
 DR
 XX
 XX Generating fusion proteins using transposable elements, useful for
 PT development of a protein domain library and in the construction of multi-
 PT functional enzymes.
 XX
 XX Example 2; Fig 2; 41pp; English.
 PS
 XX The invention relates to a Mu-like transposable element (I) used for
 CC generating functional fusion proteins after insertion into a target DNA.
 CC The Mu-like element comprises: (a) a left transposable element attachment
 CC site attL and a right transposable element attachment site attR, where
 CC attR is no more than 62 nucleotides long; (b) a site for insertion of an
 CC exogenous DNA sequence encoding for a protein domain located between attL
 CC and attR; (c) after insertion of the transposable element into a target
 CC DNA sequence, a fusion mRNA sequence is transcribed originating either
 CC from the target DNA on either side of the transposable element or from
 CC inside the transposable element and continuing through the attachment
 CC site sequences and into the protein coding region, resulting in a single
 CC fusion open reading frame (ORF). The constructs are useful in the
 CC development of a protein domain library, in the construction of multi-
 CC functional enzymes and in the accelerated evolution of new enzymatic
 CC activities. The sequences AAZ28880-228884 represent deletion mini-Mu
 CC elements of the invention (encoded ORF - AAY55901-Y55906)
 XX
 XX Sequence 71 BP; 14 A; 18 C; 16 G; 23 T; 0 U; 0 Other;
 SQ
 Query Match 78.8%; Score 39.4; DB 2; Length 71;
 Best Local Similarity 97.6%; Pred. No. 7.8e-05;
 Matches 40; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 10 GAACGAAAAACGCGAAGCGTTTCACGATAAATGCGAAAAAC 50
 DB 62 GCACGAAAAACGCGAAGCGTTTCACGATAAATGCGAAAAAC 22
 RESULT 12
 ACC80745
 ID ACC80745 standard; DNA; 82 BP.
 XX
 AC ACC80745;
 XX
 DT 15-OCT-2003 (first entry)
 XX
 XX Transposon-based targeting construct related primer Mu2-Neo-2.
 DE
 XX Targeting construct; targeting vector; transposon; recombination;
 KW

KW deletion; plant genome; animal genome; primer; PCR; ss.
 XX
 OS Synthetic.
 XX
 PN WO2003031629-A1.
 XX
 PD 17-APR-2003.
 XX
 XX 08-OCT-2002; 2002WO-AU001367.
 PF
 XX 09-OCT-2001; 2001AU-00008174.
 XX
 PR 23-MAY-2002; 2002AU-00002522.
 PR
 XX (COPY-) COPYRAT PTY LTD.
 PA
 XX Morrison J, Zhang C;
 PI
 XX WPI; 2003-393445/37.
 DR
 XX
 XX Preparing a targeting construct using a transposon and DNA recombination
 PT sequence, useful in making a targeting vector capable of modifying plant
 PT and/or animal genome in a predetermined way.
 XX
 PS Example 5; Page 51; 92pp; English.
 XX
 XX The invention relates to methods of preparing a targeting construct for
 CC use in a targeting vector capable of modifying a target DNA sequence, by
 CC obtaining a copy of the target DNA sequence in vitro, inserting a DNA
 CC sequence comprising a transposon sequence and a DNA recombination
 CC sequence at two sites in the copy of the target DNA sequence, and
 CC inducing a recombination event between the recombination sequences to
 CC delete a portion of the copy of the target DNA sequence. The methods and
 CC compositions of the present invention are useful for preparing a target
 CC construct for use in a targeting vector for gene targeting or homologous
 CC recombination. They can also be used for precisely modifying plant and/or
 CC animal genome in a predetermined way. This sequence represents a primer
 CC used in an example of the invention
 XX
 XX Sequence 82 BP; 35 A; 21 C; 16 G; 10 T; 0 U; 0 Other;
 SQ
 Query Match 78.8%; Score 39.4; DB 10; Length 82;
 Best Local Similarity 97.6%; Pred. No. 8e-05;
 Matches 40; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 10 GAACGAAAAACGCGAAGCGTTTCACGATAAATGCGAAAAAC 50
 DB 21 GCACGAAAAACGCGAAGCGTTTCACGATAAATGCGAAAAAC 61
 RESULT 13
 AAZ28881/c
 ID AAZ28881 standard; cDNA; 83 BP.
 XX
 AC AAZ28881;
 XX
 DT 15-FEB-2000 (first entry)
 XX
 XX Mini-Mu transposable element deletion region delta-66.
 DE
 XX Transposon Mu; transposable element; fusion protein; attachment site;
 KW attL; attR; protein domain library; enzyme; accelerated evolution; ss.
 XX
 OS Synthetic.
 XX
 XX US5981177-A.
 PN
 XX 09-NOV-1999.
 PD
 XX 25-JAN-1995; 95US-00378548.
 PF
 XX 25-JAN-1995; 95US-00378548.
 PR
 XX (DEMI/) DEMIRJIAN D C.
 PA

CC construct for use in a targeting vector for gene targeting or homologous
CC recombination. They can also be used for precisely modifying plant and/or
CC animal genome in a predetermined way. This sequence represents a primer
CC used in an example of the invention

XX

SQ Sequence 84 BP; 29 A; 20 C; 19 G; 16 T; 0 U; 0 Other;

Query Match 78.8%; Score 39.4; DB 10; Length 84;
Best Local Similarity 97.6%; Pred. No. 8e-05;
Matches 40; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 10 GAACGAAAAACGCGAAGCGTTTCACGATAATCGAAAAAC 50

Db 24 GCACGAAAAACGCGAAGCGTTTCACGATAATCGAAAAAC 64

Search completed: January 17, 2006, 19:35:11
Job time : 32.092 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2006 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 17, 2006, 18:09:28 ; Search time 211.625 Seconds
(without alignments)
11054.239 Million cell updates/sec

Title: US-10-511-327-5
Perfect score: 50
Sequence: 1 tgattgttgaacgaaaaac.....ttcacgataaatgcgaaaaac 50

Scoring table:
IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues ,

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_hc:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_est7:*
9: gb_gss1:*
10: gb_gss2:*
11: gb_gss3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|--------------------|
| 1 | 40.4 | 80.8 | 370 | 10 | CZ569838 OB_Ba000 |
| 2 | 39.4 | 78.8 | 226 | 9 | CC742071 ZMWB011 |
| 3 | 39.4 | 78.8 | 589 | 10 | CL858345 OR_Cha008 |
| 4 | 36.2 | 72.4 | 543 | 10 | CZ821413 OC_Ba019 |
| 5 | 36.2 | 72.4 | 688 | 10 | CZ820900 OC_Ba019 |
| 6 | 29.2 | 58.4 | 247 | 10 | CZ821171 OC_Ba019 |
| 7 | 26.6 | 53.2 | 582 | 3 | BJ016030 BJ016030 |
| 8 | 26.6 | 53.2 | 802 | 10 | DU032104 13515 Tom |
| 9 | 26.2 | 52.4 | 215 | 3 | BI953982 |
| 10 | 26.2 | 52.4 | 771 | 3 | BI953934 HVSMEM001 |
| 11 | 25.8 | 51.6 | 757 | 3 | BI953957 HVSMEM001 |
| 12 | 25.6 | 51.2 | 196 | 3 | BI953931 HVSMEM001 |
| 13 | 25.6 | 51.2 | 752 | 3 | BI953878 HVSMEM001 |
| 14 | 25.6 | 51.2 | 771 | 3 | BI953902 HVSMEM001 |
| 15 | 25.6 | 51.2 | 973 | 10 | DU005263 |
| 16 | 25.4 | 50.8 | 484 | 9 | AZ152875 SP_0046 B |
| 17 | 25.4 | 50.8 | 887 | 7 | CN586362 USDA-FP 1 |
| 18 | 25.2 | 50.4 | 274 | 8 | T31184 EST28439 Hu |
| 19 | 25.2 | 50.4 | 297 | 6 | CF504687 USDA-FP 1 |
| 20 | 25.2 | 50.4 | 491 | 8 | DN797614 USDA-FP/A |
| 21 | 25.2 | 50.4 | 737 | 3 | BI954000 HVSMEM001 |
| 22 | 25.2 | 50.4 | 741 | 3 | BI953998 HVSMEM001 |

| | | | | | | |
|---|----|------|------|-----|----|----------|
| C | 23 | 25.2 | 50.4 | 748 | 3 | BI954023 |
| | 24 | 25.2 | 50.4 | 776 | 6 | CB292852 |
| | 25 | 25.2 | 50.4 | 346 | 3 | BI507751 |
| | 26 | 24.8 | 49.6 | 705 | 3 | BQ149276 |
| | 27 | 24.6 | 49.2 | 248 | 8 | F05165 |
| | 28 | 24.6 | 49.2 | 304 | 1 | AA481614 |
| | 29 | 24.6 | 49.2 | 318 | 5 | CI4524 |
| | 30 | 24.6 | 49.2 | 403 | 1 | AA262843 |
| | 31 | 24.6 | 49.2 | 535 | 5 | CI4413 |
| | 32 | 24.6 | 49.2 | 561 | 1 | AW954984 |
| | 33 | 24.6 | 49.2 | 590 | 9 | AQ98995 |
| | 34 | 24.6 | 49.2 | 634 | 1 | AW955329 |
| | 35 | 24.6 | 49.2 | 731 | 1 | AL041260 |
| | 36 | 24.6 | 49.2 | 753 | 6 | CF449523 |
| | 37 | 24.6 | 49.2 | 766 | 3 | BI953900 |
| | 38 | 24.6 | 49.2 | 856 | 3 | BI772228 |
| | 39 | 24.6 | 49.2 | 874 | 8 | DR952125 |
| | 40 | 24.6 | 49.2 | 994 | 3 | BM804853 |
| | 41 | 24.4 | 48.8 | 592 | 9 | AZ176111 |
| | 42 | 24.4 | 48.8 | 671 | 9 | BZ833089 |
| | 43 | 24.4 | 48.8 | 703 | 5 | BW085322 |
| | 44 | 24.4 | 48.8 | 706 | 3 | BM617600 |
| | 45 | 24.4 | 48.8 | 825 | 11 | CR792741 |

ALIGNMENTS

RESULT 1
CZ569838
LOCUS
DEFINITION OB_Ba0003F02.r OB_Ba Oryza brachyantha genomic clone
OB_Ba0003F02 3', genomic survey sequence.
ACCESSION CZ569838
VERSION CZ569838.1 GI:68013639
KEYWORDS GSS.
SOURCE Oryza brachyantha
ORGANISM Oryza brachyantha
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.
REFERENCE 1 (bases 1 to 370)
Kim,H., Yu,Y., Stum,D., Yost,D., Rao,K., Luo,M., Jetty,R., Kudrna,D., Muller,C., Hatfield,J., Soderlund,C. and Wing,R.
OMAP Project
JOURNAL Unpublished (2004)
COMMENT Contact: Rod A. Wing
Arizona Genomics Institute
University of Arizona
Forbes Building Room 303, Tucson, AZ 85721-0036, USA
Tel: 520 626 9595
Fax: 520 621 1259
Email: rwing@genome.arizona.edu
PCR Primers
FORWARD: TAA TAC GAC TCA CTA TAG GG
BACKWARD: CAC TCA TTA GGC ACC CCA
Insert Length: 145 Std Error: 0.00
Plate: 0003 row: F column: 02
Seq primer: CAC TCA TTA GGC ACC CCA
Class: BAC ends.
Location/Qualifiers
1. 370
/organism="Oryza brachyantha"
/mol_type="genomic DNA"
/db_xref="taxon:4533"
/clone="OB_Ba0003F02"
/tissue_type="leaves"
/dev_stage="mature"
/lab_host="DH10B"
/clone_lib="OB_Ba"
/note="vector: PAGIBAC1; Site_1: HindIII; Site_2: HindIII"

FEATURES

source
1. 370
/organism="Oryza brachyantha"
/mol_type="genomic DNA"
/db_xref="taxon:4533"
/clone="OB_Ba0003F02"
/tissue_type="leaves"
/dev_stage="mature"
/lab_host="DH10B"
/clone_lib="OB_Ba"
/note="vector: PAGIBAC1; Site_1: HindIII; Site_2: HindIII"

ORIGIN

```

Query Match      80.8%; Score 40.4; DB 10; Length 370;
Best Local Similarity 97.6%; Pred. No. 0.00023;
Matches 41; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 9 TGAACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAC 50
   ||||||||||||||||||||||||||||||||||||||||
Db 85 TGCAGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAC 126

RESULT 2
CC742071
LOCUS CC742071
DEFINITION ZMMBB0115J07.f ZMMBB Zea mays genomic clone ZMMBB0115J07 5',
genomic survey sequence.
ACCESSION CC742071
VERSION CC742071.1 GI:32194524
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 226)
AUTHORS Yu, Y., Kim, H.R., Hatfield, J., Soderlund, C., Bharti, A.K., Messing, J.
and Wing, R.
TITLE Sequencing of the maize genome
JOURNAL Unpublished (2003)
COMMENT Contact: Rod Wing
Arizona Genomics Institute
University of Arizona
Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ
85721-0088, USA
Tel: 520 626 3967
Fax: 520 621 9288
Email: rwing@genome.arizona.edu
PCR Primers
FORWARD: M13r
BACKWARD: M13r
Plate: 0115 row: J column: 07
Seq primer: T7
Class: BAC ends.
FEATURES
source
Location/Qualifiers
1..226
/organism="Zea mays"
/mol_type="genomic DNA"
/cultivar="B73"
/db_xref="taxon:4577"
/clone="ZMMBB0115J07"
/lab_host="DH10B"
/clone_lib="ZMMBB"
/note="Vector: pBelobAC11; Site_1: HindIII; Site_2:
HindIII; Zea mays L. ssp. mays"

ORIGIN
Query Match      78.8%; Score 39.4; DB 9; Length 226;
Best Local Similarity 97.6%; Pred. No. 0.00053;
Matches 40; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 10 GAACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAC 50
   ||||||||||||||||||||||||||||||||||||||||
Db 182 GCACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAC 222

RESULT 3
CL858345
LOCUS CL858345
DEFINITION OR_CBa0089D12.f OR_CBa Oryza rufipogon genomic clone OR_CBa0089D12
5', genomic survey sequence.
ACCESSION CL858345
VERSION CL858345.1 GI:51269584
KEYWORDS GSS.
SOURCE Oryza rufipogon
ORGANISM Oryza rufipogon
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.
1 (bases 1 to 589)
Kim, H., Yu, Y., Wissotski, M., Yost, D., Stum, D., Rao, K., Luo, M.,
Jetty, R., Kudrna, D., Muller, C., Hatfield, J., Soderlund, C. and
Wing, R.
OMAP project
Unpublished (2004)
Contact: Rod A. Wing
Arizona Genomics Institute
University of Arizona
Forbes Building Room 303, Tucson, AZ 85721-0036, USA
Tel: 520 626 9595
Fax: 520 621 1259
Email: rwing@genome.arizona.edu
PCR Primers
FORWARD: TAA TAC GAC TCA CTA TAG GG
BACKWARD: CAC TCA TTA GGC ACC CCA
Plate: 0089 row: D column: 12
Seq primer: TAA TAC GAC TCA CTA TAG GG
Class: BAC ends.
FEATURES
source
Location/Qualifiers
1..589
/organism="Oryza rufipogon"
/mol_type="genomic DNA"
/db_xref="taxon:4529"
/clone="OR_CBa0089D12"
/tissue_type="young leaves"
/dev_stage="2 week old seedlings"
/lab_host="DH10B T1 phage resistant"
/clone_lib="OR CBa"
/note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII;
drk treated 36 hrs before harvest"

ORIGIN
Query Match      78.8%; Score 39.4; DB 10; Length 589;
Best Local Similarity 97.6%; Pred. No. 0.00054;
Matches 40; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 10 GAACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAC 50
   ||||||||||||||||||||||||||||||||||||||||
Db 495 GCACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAC 535

RESULT 4
CZ821413
LOCUS CZ821413
DEFINITION OC_Ba0199p13.r OC_Ba Oryza coarctata genomic clone OC_Ba0199p13
3', genomic survey sequence.
ACCESSION CZ821413
VERSION CZ821413.1 GI:71261266
KEYWORDS GSS.
SOURCE Oryza coarctata (Porteresia coarctata)
ORGANISM Oryza coarctata
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.
1 (bases 1 to 543)
Kim, H., Collura, K., Wissotski, M., Byrne, M., Stum, D., Smart, D.,
Rao, K., Luo, M., Jetty, R., Kudrna, D., Muller, C., Soderlund, C. and
Wing, R.
OMAP (Oryza Map Alignment Project)- Arizona Genomics Institute
Unpublished (2005)
Contact: Rod A. Wing
Arizona Genomics Institute
University of Arizona
Forbes Building Room 303, Tucson, AZ 85721-0036, USA
Tel: 520 626 9595
Fax: 520 621 1259
Email: rwing@genome.arizona.edu
PCR Primers
FORWARD: TAA TAC GAC TCA CTA TAG GG

```

```

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.
1 (bases 1 to 589)
Kim, H., Yu, Y., Wissotski, M., Yost, D., Stum, D., Rao, K., Luo, M.,
Jetty, R., Kudrna, D., Muller, C., Hatfield, J., Soderlund, C. and
Wing, R.
OMAP project
Unpublished (2004)
Contact: Rod A. Wing
Arizona Genomics Institute
University of Arizona
Forbes Building Room 303, Tucson, AZ 85721-0036, USA
Tel: 520 626 9595
Fax: 520 621 1259
Email: rwing@genome.arizona.edu
PCR Primers
FORWARD: TAA TAC GAC TCA CTA TAG GG
BACKWARD: CAC TCA TTA GGC ACC CCA
Plate: 0089 row: D column: 12
Seq primer: TAA TAC GAC TCA CTA TAG GG
Class: BAC ends.
FEATURES
source
Location/Qualifiers
1..589
/organism="Oryza rufipogon"
/mol_type="genomic DNA"
/db_xref="taxon:4529"
/clone="OR_CBa0089D12"
/tissue_type="young leaves"
/dev_stage="2 week old seedlings"
/lab_host="DH10B T1 phage resistant"
/clone_lib="OR CBa"
/note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII;
drk treated 36 hrs before harvest"

ORIGIN
Query Match      78.8%; Score 39.4; DB 10; Length 589;
Best Local Similarity 97.6%; Pred. No. 0.00054;
Matches 40; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 10 GAACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAC 50
   ||||||||||||||||||||||||||||||||||||||||
Db 495 GCACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAC 535

RESULT 4
CZ821413
LOCUS CZ821413
DEFINITION OC_Ba0199p13.r OC_Ba Oryza coarctata genomic clone OC_Ba0199p13
3', genomic survey sequence.
ACCESSION CZ821413
VERSION CZ821413.1 GI:71261266
KEYWORDS GSS.
SOURCE Oryza coarctata (Porteresia coarctata)
ORGANISM Oryza coarctata
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.
1 (bases 1 to 543)
Kim, H., Collura, K., Wissotski, M., Byrne, M., Stum, D., Smart, D.,
Rao, K., Luo, M., Jetty, R., Kudrna, D., Muller, C., Soderlund, C. and
Wing, R.
OMAP (Oryza Map Alignment Project)- Arizona Genomics Institute
Unpublished (2005)
Contact: Rod A. Wing
Arizona Genomics Institute
University of Arizona
Forbes Building Room 303, Tucson, AZ 85721-0036, USA
Tel: 520 626 9595
Fax: 520 621 1259
Email: rwing@genome.arizona.edu
PCR Primers
FORWARD: TAA TAC GAC TCA CTA TAG GG

```

BACKWARD: CAC TCA TTA GGC ACC CCA
 Plate: 0199 row: P column: 13
 Seq primer: CAC TCA TTA GGC ACC CCA
 Class: BAC ends.

FEATURES

Location/Qualifiers
 1..543
 /organism="Oryza coarctata"
 /mol_type="genomic DNA"
 /db_xref="taxon:77588"
 /clone="OC_Ba0199p13"
 /tissue_type="leaves"
 /dev_stage="mature"
 /lab_host="DH10B"
 /clone_lib="OC_Ba"
 /note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII"

ORIGIN

Query Match 72.4%; Score 36.2; DB 10; Length 543;
 Best Local Similarity 92.7%; Pred. No. 0.0083;
 Matches 38; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 10 GAACGAAACCGGAAAGCGTTTCACGATAAATGCGAAAC 50
 |
 Db 473 GCACGAAATCGGAAAGCGTTTCACGATACATGCGAAAC 513
 |

RESULT 5
 C2820900 688 bp DNA linear GSS 26-JUL-2005
 LOCUS OC_Ba0199D21.r OC_Ba Oryza coarctata genomic clone OC_Ba0199D21
 DEFINITION 3', genomic survey sequence.

ACCESSION

VERSION C2820900.1 GI:71260753

KEYWORDS

SOURCE Oryza coarctata (Porteresia coarctata)

ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Ehrhartoideae; Oryzeae; Oryza.

REFERENCE

1 (bases 1 to 688)
 Kim, H., Collura, K., Wisotski, M., Byrne, M., Stum, D., Smart, D.,
 Rao, K., Luo, M., Jetty, R., Kudrna, D., Muller, C., Soderlund, C. and
 Wing, R.

AUTHORS

OMAP (Oryza Map Alignment Project)- Arizona Genomics Institute

TITLE

Unpublished (2005)

JOURNAL

COMMENT Arizona Genomics Institute

University of Arizona

Forbes Building Room 303, Tucson, AZ 85721-0036, USA

Tel: 520 626 9595

Fax: 520 621 1259

Email: rwing@genome.arizona.edu

PCR Primers

FORWARD: TAA TAC GAC TCA CTA TAG GG

BACKWARD: CAC TCA TTA GGC ACC CCA

Plate: 0199 row: D column: 21

Seq primer: CAC TCA TTA GGC ACC CCA

Class: BAC ends.

FEATURES

Location/Qualifiers
 1..688
 /organism="Oryza coarctata"
 /mol_type="genomic DNA"
 /db_xref="taxon:77588"
 /clone="OC_Ba0199D21"
 /tissue_type="leaves"
 /dev_stage="mature"
 /lab_host="DH10B"
 /clone_lib="OC_Ba"
 /note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII"

ORIGIN

Query Match 72.4%; Score 36.2; DB 10; Length 688;
 Best Local Similarity 92.7%; Pred. No. 0.0083;

Matches 38; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 10 GAACGAAACCGGAAAGCGTTTCACGATAAATGCGAAAC 50
 |
 Db 440 GCACGAAACCGGACAGCGTTCACGATAAATGCGAAAC 480
 |

RESULT 6

C2821171

LOCUS

DEFINITION

3', genomic survey sequence.

ACCESSION

VERSION C2821171.1 GI:71261024

KEYWORDS

SOURCE

ORGANISM

Oryza coarctata

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Ehrhartoideae; Oryzeae; Oryza.

REFERENCE

1 (bases 1 to 247)

AUTHORS

Kim, H., Collura, K., Wisotski, M., Byrne, M., Stum, D., Smart, D.,
 Rao, K., Luo, M., Jetty, R., Kudrna, D., Muller, C., Soderlund, C. and
 Wing, R.

OMAP (Oryza Map Alignment Project)- Arizona Genomics Institute

Unpublished (2005)

JOURNAL

COMMENT Arizona Genomics Institute

University of Arizona

Forbes Building Room 303, Tucson, AZ 85721-0036, USA

Tel: 520 626 9595

Fax: 520 621 1259

Email: rwing@genome.arizona.edu

PCR Primers

FORWARD: TAA TAC GAC TCA CTA TAG GG

BACKWARD: CAC TCA TTA GGC ACC CCA

Plate: 0199 row: J column: 21

Seq primer: CAC TCA TTA GGC ACC CCA

Class: BAC ends.

FEATURES

Location/Qualifiers

1..247

/organism="Oryza coarctata"

/mol_type="genomic DNA"

/db_xref="taxon:77588"

/clone="OC_Ba0199J21"

/tissue_type="leaves"

/dev_stage="mature"

/lab_host="DH10B"

/clone_lib="OC_Ba"

/note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII"

ORIGIN

Query Match 58.4%; Score 29.2; DB 10; Length 247;

Best Local Similarity 81.0%; Pred. No. 3.2;

Matches 34; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 9 TGAACGAAACCGGAAAGCGTTTCACGATAAATGCGAAAC 50
 |
 Db 143 TGCATCAAAATCGAAAGCGTTTCATGATCAATGTGAAAC 184
 |

RESULT 7

BJ016030

LOCUS

DEFINITION

mRNA sequence.

ACCESSION

VERSION BJ016030.1 GI:17375178

KEYWORDS

SOURCE

ORGANISM

Oryzias latipes

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Euteleostei;

Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
Belontiiformes; Adrianichthyidae; Oryziinae; Oryzias.

1 (bases 1 to 592)
Kohara,Y., Shin-I,T., Kimura,T., Narita,T., Jindo,T. and Takeda,H.
Medaka EST Project in Takeda's lab
Unpublished (2001)
Contact: Tadaeu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.

FEATURES

Location/Qualifiers

1..582

/organism="Oryzias latipes"
/mol_type="mRNA"
/strains="Hd-rR"
/db_xref="taxon:8090"
/clone="MF01SSA008A02"
/sex="mixture of female and male"
/tissue_type="whole embryo"
/dev_stage="segmentation stage 20 - 25"
/clone_lib="MF01SSA cDNA"

ORIGIN

Query Match 53.2%; Score 26.6; DB 3; Length 582;
Best Local Similarity 71.4%; Pred. No. 30;
Matches 35; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

Oy 1 TCATTGTTGACGAAACCGGAAGCGTTTCAGATAAATGCCAAAA 49

Db 210 TCATTTATTAAACGCGAAACCAAAAGCATTTTCAGATAAATCACTAACA 258

RESULT 8

DU032104/c

LOCUS DU032104 802 bp DNA linear GSS 12-AUG-2005
DEFINITION 13515 Tomato HindIII BAC Library Lycopersicon esculentum genomic
clone LE_HBa0169A20 3, genomic survey sequence.

ACCESSION DU032104

VERSION DU032104.1 GI:72445405

KEYWORDS GSS.

SOURCE Lycopersicon esculentum (Solanum lycopersicum)

ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
asterids; lamids; Solanales; Solanaceae; Solanum; Lycopersicon.

REFERENCE

AUTHORS Mueller,L.A., Buels,R.M., Wang,Y., Tanksley,S.D., Giovannoni,J.J.,
Van Eck,J. and Stack,S.
BAC end sequencing from three Solanum lycopersicon libraries
Unpublished (2005)

TITLE

JOURNAL

COMMENT

Contact: Lukas Mueller
Tanksley Lab, Dept. of Plant Breeding
Cornell University
251 Emerson Hall, Ithaca, NY 14853, USA
Tel: 607-255-6557
Fax: 607-255-6683
Email: sgn-feedback@cornell.edu
Plate: 169 row: A column: 20

Seq primer: Sp6

Class: BAC ends

High quality sequence start: 36

High quality sequence stop: 428.

Location/Qualifiers

1..802

/organism="Lycopersicon esculentum"
/mol_type="Genomic DNA"
/cultivar="Heinz 1706"
/db_xref="taxon:4081"
/clone="LE_HBa0169A20"

FEATURES

source

ORIGIN

Query Match 53.2%; Score 26.6; DB 10; Length 802;
Best Local Similarity 71.4%; Pred. No. 30;
Matches 35; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

Oy 2 GATTGATTGAACGAAAAACGCGAAGCGTTTCAGATAAATGCCAAAAAC 50

Db 693 GATGATTAAAGAAAAAGCAAGATTTTCACTAAATATGAGAAAAAC 645

RESULT 9

BI953982/c

LOCUS

DEFINITION BI953982 215 bp mRNA linear EST 19-OCT-2001
HVSME0015K12f Hordeum vulgare green seedling EST library
HVCDNA0014 (Blumeria infected) Hordeum vulgare subsp. vulgare cDNA
clone HVSME0015K12f, mRNA sequence.

ACCESSION BI953982

VERSION BI953982.1

KEYWORDS EST.

SOURCE

ORGANISM Hordeum vulgare subsp. vulgare

Hordeum vulgare subsp. vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Poideae; Triticeae; Hordeum.

REFERENCE 1 (bases 1 to 215)

AUTHORS Wing,R., Close,T.J., Kleinhofs,A., Wise,R., Chin,A., Begum,D.,
Frisch,D., Atkins,M., Yu,Y., Henry,D., Palmer,M., Rambo,T.,
Simmons,J., Oates,R. and Main,D.

TITLE Development of a genetically and physically anchored EST resource
for barley genomics: Blumeria infected Morex (compatible) seedling
cDNA library

JOURNAL

COMMENT Unpublished (2001)

Contact: Wing RA

Clemson University Genomics Institute

Clemson University

100 Jordan Hall, Clemson, SC 29634, USA

Tel: 864 656 7288

Fax: 864 656 4293

Email: rwing@clemson.edu

Total hg bases = 174

Seq primer: AATTAACCTCACTAAAGG

High quality sequence start: 4

High quality sequence stop: 216.

Location/Qualifiers

1..215

/organism="Hordeum vulgare subsp. vulgare"

/mol_type="mRNA"

/cultivar="Morex"

/sub_species="vulgare"

/db_xref="taxon:112509"

/clone="HVSME0015K12f"

/tissue_type="green seedling leaf"

/lab_host="TJCl21"

/clone_lib="Hordeum vulgare green seedling EST library
HVCDNA0014 (Blumeria infected)"

/note="vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI; Morex (mla) plants were greenhouse grown in the R
Wise lab at Iowa State University, Ames, IA; 7 day old
green seedlings were infected with isolate 5874 of
Blumeria graminis f. sp. hordei, and leaves were harvested

24, 48 and 72 hr post-inoculation and snap frozen (Wise).
In the TJ Close lab at the University of California,
Riverside, total RNA was prepared from each sample pool,
equal quantities of all three RNA pools were combined,

poly(A) RNA was purified from the mixture, one primary
unamplified cDNA library was made, and 1 million pfu were
in vivo excised to give pBluescript SK(-) cDNA phagemids
(Chin). Phagemids were plated and picked at the Clemson

University Genomics Institute (CUGI) (Begum, Palmer,
University Genomics Institute (CUGI) (Begum, Palmer,


```

/db_xref="taxon:112509"
/clone="HVSME0015112f"
/tissue_type="green seedling leaf"
/lab_host="TJCl21"
/clone_lib="Hordeum vulgare green seedling EST library
HVCdNA0014 (Blumeria infected)"
/notes="vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI; Morex (mla) plants were greenhouse grown in the R
Wise lab at Iowa State University, Ames, IA; 7 day old
green seedlings were infected with isolate 5874 of
Blumeria graminis f. sp. hordei, and leaves were harvested
24, 48 and 72 hr post-inoculation and snap frozen (Wise).
In the TJ Close lab at the University of California,
Riverside, total RNA was prepared from each sample pool,
equal quantities of all three RNA pools were combined,
poly(A) RNA was purified from the mixture, one primary
unamplified cDNA library was made, and 1 million pfu were
in vivo excised to give pBluescript SK(-) cDNA phagemids
(Chin). Phagemids were plated and picked at the Clemson
University Genomics Institute (CUGI) (Begum, Palmer,
Frisch, Atkins and Wing). Plasmid DNA preparations, DNA
sequencing and sequence analysis were performed at CUGI
(Wing, Yu, Frisch, Henry, Simmons, Oates, Rambo, Main).
The sequence has been trimmed to remove vector sequence
and contains a minimum of 100 bases of phred value 20 or
above. For more details on library preparation and
sequence analysis see
http://www.genome.clemson.edu/projects/barley. To order
this clone see http://www.genome.clemson.edu/orders_Also
see Close TJ, Wing R, Kleinhofs A, Wise R (2001)
Genetically and physically anchored EST resources for
barley genomics. Barley Genetics Newsletter 31:29-30.
(http://wheat.pw.usda.gov/ggp/bgn/31/cover.html)"

```

ORIGIN

```

Query Match      51.6%; Score 25.8; DB 3; Length 757;
Best Local Similarity 92.7%; Pred. No. 60;
Matches 38; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

Qy 10 GAACGAAAAACCGAAAGCGTTTCACGATAATGCGAAAC 50
    |
Db 40 GCACGAAAAACCGAAAGCGTTTCACG-TAAATGCGAAAC 1

```

RESULT 12

```

BI953931/c
LOCUS
DEFINITION
HVCdNA0014 (Blumeria infected) Hordeum vulgare subsp. vulgare cDNA
clone HVSME0015112f, mRNA sequence.

```

ACCESSION

```

VERSION
BI953931.1 GI:16299011

```

KEYWORDS

```

SOURCE

```

ORGANISM

```

Hordeum vulgare subsp. vulgare
Hordeum vulgare subsp. vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Poideae; Triticeae; Hordeum.

```

REFERENCE

```

AUTHORS
Wing,R., Close,T.J., Kleinhofs,A., Wise,R., Chin,A., Begum,D.,
Frisch,D., Atkins,M., Yu,Y., Henry,D., Palmer,M., Rambo,T.,
Simmons,J., Oates,R. and Main,D.
Development of a genetically and physically anchored EST resource
for barley genomics: Blumeria infected Morex (compatible) seedling
cDNA library
Unpublished (2001)
Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu

```

TITLE

JOURNAL

COMMENT

FEATURES

source

```

1. .196
/organism="Hordeum vulgare subsp. vulgare"
/mol_type="mRNA"
/cultivar="Morex"
/sub_species="vulgare"
/db_xref="taxon:112509"
/clone="HVSME0015112f"
/tissue_type="green seedling leaf"
/lab_host="TJCl21"
/clone_lib="Hordeum vulgare green seedling EST library
HVCdNA0014 (Blumeria infected)"
/notes="vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI; Morex (mla) plants were greenhouse grown in the R
Wise lab at Iowa State University, Ames, IA; 7 day old
green seedlings were infected with isolate 5874 of
Blumeria graminis f. sp. hordei, and leaves were harvested
24, 48 and 72 hr post-inoculation and snap frozen (Wise).
In the TJ Close lab at the University of California,
Riverside, total RNA was prepared from each sample pool,
equal quantities of all three RNA pools were combined,
poly(A) RNA was purified from the mixture, one primary
unamplified cDNA library was made, and 1 million pfu were
in vivo excised to give pBluescript SK(-) cDNA phagemids
(Chin). Phagemids were plated and picked at the Clemson
University Genomics Institute (CUGI) (Begum, Palmer,
Frisch, Atkins and Wing). Plasmid DNA preparations, DNA
sequencing and sequence analysis were performed at CUGI
(Wing, Yu, Frisch, Henry, Simmons, Oates, Rambo, Main).
The sequence has been trimmed to remove vector sequence
and contains a minimum of 100 bases of phred value 20 or
above. For more details on library preparation and
sequence analysis see
http://www.genome.clemson.edu/projects/barley. To order
this clone see http://www.genome.clemson.edu/orders_Also
see Close TJ, Wing R, Kleinhofs A, Wise R (2001)
Genetically and physically anchored EST resources for
barley genomics. Barley Genetics Newsletter 31:29-30.
(http://wheat.pw.usda.gov/ggp/bgn/31/cover.html)"

```

ORIGIN

```

Query Match      51.2%; Score 25.6; DB 3; Length 196;
Best Local Similarity 87.5%; Pred. No. 69;
Matches 28; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

```

Qy 10 GAACGAAAAACCGAAAGCGTTTCACGATAAA 41
    |
Db 32 GCACGAAAAACCGAAAGCGTTTCACGTAATA 1

```

RESULT 13

BI953878/c

LOCUS

```

DEFINITION
HVSME0015112f Hordeum vulgare green seedling EST library
clone HVSME0015112f, mRNA sequence.

```

ACCESSION

```

VERSION
BI953878.1 GI:16298915

```

KEYWORDS

```

SOURCE

```

ORGANISM

```

Hordeum vulgare subsp. vulgare
Hordeum vulgare subsp. vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Poideae; Triticeae; Hordeum.

```

REFERENCE

```

AUTHORS
Wing,R., Close,T.J., Kleinhofs,A., Wise,R., Chin,A., Begum,D.,
Frisch,D., Atkins,M., Yu,Y., Henry,D., Palmer,M., Rambo,T.,
Simmons,J., Oates,R. and Main,D.
Development of a genetically and physically anchored EST resource

```

TITLE

GenCore version 5.1.6
Copyright (c) 1993 - 2006 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 17, 2006, 18:06:53 ; Search time 24804 Seconds
(without alignments)
11032.288 Million cell updates/sec

Title: US-10-511-327-7
Perfect score: 4814
Sequence: 1 ggtaccctggaatgcgcaa.....taagatcgtttctgtgact 4814

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl.*

1: gb_ba.*
2: gb_in.*
3: gb_env.*
4: gb_on.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pr.*
9: gb_ro.*
10: gb_sts.*
11: gb_sy.*
12: gb_un.*
13: gb_vi.*
14: gb_htg.*
15: gb_pl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|---------------|-------------|
| 1 | 1636.8 | 34.0 | 15129 | 11 AY744149 | Dengue vi |
| 2 | 1636.8 | 34.0 | 15129 | 11 AY744150 | Dengue vi |
| 3 | 1636.8 | 34.0 | 15145 | 11 AY656170 | Dengue vi |
| 4 | 1636.8 | 34.0 | 15159 | 11 AY744148 | Dengue vi |
| 5 | 1636.8 | 34.0 | 15176 | 11 AY656169 | Dengue vi |
| 6 | 1636.8 | 34.0 | 15237 | 11 AY243469 | Chimeric |
| 7 | 1636.8 | 34.0 | 15239 | 11 AY243467 | Chimeric |
| 8 | 1636.8 | 34.0 | 15239 | 11 AY376438 | Dengue vi |
| 9 | 1636.8 | 34.0 | 15256 | 11 AY656168 | Chimeric |
| 10 | 1636.8 | 34.0 | 15268 | 11 AY243468 | Chimeric |
| 11 | 1636.8 | 34.0 | 15270 | 11 AY243466 | Chimeric |
| 12 | 1636.8 | 34.0 | 15270 | 11 AY648301 | Dengue vi |
| 13 | 1636.8 | 34.0 | 15287 | 11 AY656167 | Chimeric |
| 14 | 1635.8 | 34.0 | 13561 | 11 AY028776 | TnpHoZ mu |
| 15 | 1624 | 33.7 | 6343 | 11 SYNCLCTAPS | Expression |
| 16 | 1624 | 33.7 | 6571 | 11 SYNCLCTAPA | Expression |
| 17 | 1624 | 33.7 | 8071 | 11 SYNCLCTAP | Expression |
| 18 | 1560.6 | 32.4 | 7311 | 6 CS091381 | Sequence |

| | | | | | |
|------|--------|------|-------|--------------|--------------------|
| C 19 | 1469 | 30.5 | 5228 | 11 XXU25059 | U25059 Cloning vec |
| C 20 | 1452 | 30.2 | 56167 | 3 AU635924 | AU635924 Plasmid p |
| C 21 | 1450.4 | 30.1 | 11823 | 1 AY043299 | AY043299 Aeromonas |
| C 22 | 1450.4 | 30.1 | 11943 | 1 SYNPRF339 | D45834 Size marker |
| C 23 | 1448.2 | 30.1 | 3779 | 11 SYNPRF339 | L09155 pWT571 expr |
| C 24 | 1447.8 | 30.1 | 4840 | 6 AX084356 | AX084356 Sequence |
| C 25 | 1447.8 | 30.1 | 5817 | 11 YRP7 | U03501 Yeast repli |
| C 26 | 1447.8 | 30.1 | 10667 | 11 YEP13 | U03498 Yeast epis |
| C 27 | 1447.8 | 30.1 | 10667 | 11 YEP213 | U03499 Yeast epis |
| C 28 | 1447 | 30.1 | 6303 | 11 AY093430 | AY093430 Allelic e |
| C 29 | 1447 | 30.1 | 6760 | 11 AY093429 | AY093429 Allelic e |
| C 30 | 1446.8 | 30.1 | 3423 | 6 AR287564 | AR287564 Sequence |
| C 31 | 1446.8 | 30.1 | 3474 | 6 A28084 | A28084 pMTNF-MPH p |
| C 32 | 1446.8 | 30.1 | 3474 | 6 A28085 | A28085 pMTNF-MPH p |
| C 33 | 1446.8 | 30.1 | 3474 | 6 A75741 | A75741 Sequence 10 |
| C 34 | 1446.8 | 30.1 | 3474 | 6 AR085100 | AR085100 Sequence |
| C 35 | 1446.8 | 30.1 | 3474 | 6 AR209770 | AR209770 Sequence |
| C 36 | 1446.8 | 30.1 | 3474 | 6 AR287565 | AR287565 Sequence |
| C 37 | 1446.8 | 30.1 | 3801 | 6 AR493833 | AR493833 Sequence |
| C 38 | 1446.8 | 30.1 | 4245 | 6 AR069365 | AR069365 Sequence |
| C 39 | 1446.8 | 30.1 | 4245 | 11 PACYC184 | X06403 Cloning vec |
| C 40 | 1446.8 | 30.1 | 4633 | 6 A20355 | A20355 plasmid pIG |
| C 41 | 1446.8 | 30.1 | 4633 | 6 A20356 | A20356 plasmid pIG |
| C 42 | 1446.8 | 30.1 | 4753 | 11 AF129432 | AF129432 Cloning v |
| C 43 | 1446.8 | 30.1 | 5201 | 6 AR493834 | AR493834 Sequence |
| C 44 | 1446.8 | 30.1 | 5281 | 11 SCU22104 | U22104 Cloning vec |
| C 45 | 1446.8 | 30.1 | 5443 | 11 AY796342 | AY796342 Cloning v |

ALIGNMENTS

RESULT 1
AY744149
LOCUS AY744149 15129 bp DNA circular SYN 03-DEC-2004
DEFINITION Dengue virus type 2 vector p2(delta30), complete sequence.
ACCESSION AY744149
VERSION AY744149.1 GI:56089726
KEYWORDS
SOURCE
ORGANISM
REFERENCE 1 (bases 1 to 15129)
AUTHORS Blaney, J.E. Jr., Hanson, C.T., Hanley, K.A., Murphy, B.R. and Whitehead, S.S.
TITLE Vaccine candidates derived from a novel infectious cDNA clone of an American genotype dengue virus type 2
JOURNAL (Er) BMC Infect. Dis. 4 (1), 39 (2004)
PUBMED 15461822
REFERENCE 2 (bases 1 to 15129)
AUTHORS Whitehead, S.S.
TITLE Direct Submission
JOURNAL Submitted (07-SEP-2004) Laboratory of Infectious Diseases, NIAID, 50 South Drive, Bldg 50, Room 6515, Bethesda, MD 20892, USA
FEATURES
source
1. .15129
/organism="Dengue virus type 2 vector p2(delta30)"
/mol_type="other DNA"
/strain="Tonga/74"
/db_xref="taxon:301536"
/country="Tonga"
/focus
1. .10683
/organism="Dengue virus type 2"
/mol_type="other DNA"
/db_xref="taxon:11060"
10593. .15054
/organism="Cloning vector pBR322"
/mol_type="other DNA"
/db_xref="taxon:47470"
97. .10272
/codon_start=1
/product="polyprotein"


```
Db 11942 TTTGTTCCGGTGGGATATGGTGGAGGCCCCGGTGGCGGAGCATGTTGGGGCGCATCTC 12001
Qy 1889 CTTGCATGACCAATTCCTTGGGGGGGGTGTCTCAACGGGCTCTCAACCTACTACTACTGGGCTG 1948
Db 12002 CTTGCATGACCAATTCCTTGGGGGGGGTGTCTCAACGGGCTCTCAACCTACTACTACTGGGCTG 12061
Qy 1949 CTTCTTAATGAGAGTGCATATAGGAGAGAGGTGCAGCATGCGCTTGGAGAGCTTCAA 2008
Db 12062 CTTCTTAATGAGAGTGCATATAGGAGAGAGGTGCAGCATGCGCTTGGAGAGCTTCAA 12121
Qy 2009 CCCAGTCACTCTTCCGGTGGGGGGGGGAGATGACTATCTGTCGCGCACCTTATGACTGT 2068
Db 12122 CCCAGTCACTCTTCCGGTGGGGGGGGGAGATGACTATCTGTCGCGCACCTTATGACTGT 12181
Qy 2069 CTTCTTTATCATCAACTCGTAGGACAGGTGCGGCGAGCGTCTGGGTCACTTTTCGGGGA 2128
Db 12182 CTTCTTTATCATCAACTCGTAGGACAGGTGCGGCGAGCGTCTGGGTCACTTTTCGGGGA 12241
Qy 2129 GGACCGCTTTCGCTGGAGCGGACGATGATCGGCTGTGCTTGGGTGATTTTCGGAATCTT 2188
Db 12242 GGACCGCTTTCGCTGGAGCGGACGATGATCGGCTGTGCTTGGGTGATTTTCGGAATCTT 12301
Qy 2189 GCAGCGCTCTCAAGCCTTCTGCTCACTGTCCTCGGCGACCAAAAGTTTCGGGGAGAGCA 2248
Db 12302 GCAGCGCTCTCAAGCCTTCTGCTCACTGTCCTCGGCGACCAAAAGTTTCGGGGAGAGCA 12361
Qy 2249 GGCAATTATCCCGGCGATGGCGGCGACGCGCTGGGCTACGCTTCTGCTGGGTTTCGGGAC 2308
Db 12362 GGCAATTATCCCGGCGATGGCGGCGACGCGCTGGGCTACGCTTCTGCTGGGTTTCGGGAC 12421
Qy 2309 GCGAGGTGGATGCGCTTCCCATTTATGATTTCTTCTGCTTCCGGCGGACATCGGGATGCC 2368
Db 12422 GCGAGGTGGATGCGCTTCCCATTTATGATTTCTTCTGCTTCCGGCGGACATCGGGATGCC 12481
Qy 2369 GCGCTTGCAGCGCATGCTGTCCAGCGAGTATGATGACGACCATCAGGACAGCTTCAAGG 2428
Db 12482 GCGCTTGCAGCGCATGCTGTCCAGCGAGTATGATGACGACCATCAGGACAGCTTCAAGG 12541
Qy 2429 ATCGCTCGCGCTCTTACACGCTTAACCTTTCATGATTTGACCGCTGATCGTCACGGCGAT 2488
Db 12542 ATCGCTCGCGCTCTTACACGCTTAACCTTTCATGATTTGACCGCTGATCGTCACGGCGAT 12601
Qy 2489 TTATGCCGCTTCGCGAGCACATGGAACGGGTGGCATGATTTAGCGCGCGCCCTATA 2548
Db 12602 TTATGCCGCTTCGCGAGCACATGGAACGGGTGGCATGATTTAGCGCGCGCCCTATA 12661
Qy 2549 CTTTGTCTGCCCTCCCGGTTGCGTCCGGGTGCGATGAGCGCGGCGACCTCGACTGAAT 2608
Db 12662 CTTTGTCTGCCCTCCCGGTTGCGTCCGGGTGCGATGAGCGCGGCGACCTCGACTGAAT 12721
Qy 2609 GGAAGCGGGCGCACCTCGCTAACGGATTACACCTTCCAGAAATTTGGAGCCAAATCAATTC 2668
Db 12722 GGAAGCGGGCGCACCTCGCTAACGGATTACACCTTCCAGAAATTTGGAGCCAAATCAATTC 12781
Qy 2669 TTGCGGAGAACTGTGAATGCGCAAAACCAACCTTTGGCAGAACATATCATCGCTCGGCG 2728
Db 12782 TTGCGGAGAACTGTGAATGCGCAAAACCAACCTTTGGCAGAACATATCATCGCTCGGCG 12841
Qy 2729 ATCTCCAGAGCGGACGCGCGCATCTCTGGGAGCGGTGGGTCTCTGG----- 2776
Db 12842 ATCTCCAGAGCGGACGCGCGCATCTCTGGGAGCGGTGGGTCTCTGGCCACGGGTGCGC 12901
Qy 2777 ----- 2776
Db 12902 ATGATCGTGTCTCTGTGTTGAGACCGCGGTAGGCTGGGGGTGCTTACTGGTTAG 12961
Qy 2777 ----- 2776
Db 12962 CAGAAATGAAATCACCAGTACGGAGCGGAGCACTGAAGGAGCTGCTGCTGCAAAACGCTCTGG 13021
Qy 2777 ----- 2776
```

```
Db 13022 ACCTGAGCAACAACATGAATGGTCTTCGGTTTCCGTGTTTCGTAAGTCTCGAAACGCGG 13081
Qy 2777 ----- 2776
Db 13082 AAGTCAGCGCCTGCACCATTTATGTTCCGGATCTGCATCGCAGGATGCTGCTGGCTACCC 13141
Qy 2777 -----GCTGGCATTTGACCTGAGTGAATTTTCTC 2805
Db 13142 TGTGAAACACCTACATCTGTATTAACGAAGCGCTGGCATTTGACCTGAGTGAATTTTCTC 13201
Qy 2806 TGGTCCCGCGCATCTCAATACCGCAGTTGTTTACCTCACAACGTTCCAGTAACCGGCA 2865
Db 13202 TGGTCCCGCGCATCTCAATACCGCAGTTGTTTACCTCACAACGTTCCAGTAACCGGCA 13261
Qy 2866 TGTTCATCATCAGTAACCCGCTATCGTGAGCATCTCTCTCGTTTTCATCGGTATCATTTACC 2925
Db 13262 TGTTCATCATCAGTAACCCGCTATCGTGAGCATCTCTCTCGTTTTCATCGGTATCATTTACC 13321
Qy 2926 CCGATGAACAGAAATCCCTTTACACGGAGCATCAGTGACCAACAGGAAAAAACCGCC 2985
Db 13322 CCGATGAACAGAAATCCCTTTACACGGAGCATCAGTGACCAACAGGAAAAAACCGCC 13381
Qy 2986 CTTAACATGCGCCGCTTTATCAGAAGCCAGACATTAACGTTCTTGGAGAACTCAACGAG 3045
Db 13382 CTTAACATGCGCCGCTTTATCAGAAGCCAGACATTAACGTTCTTGGAGAACTCAACGAG 13441
Qy 3046 CTGACCGCGATGAACAGGACAGCTCT 3073
Db 13442 CTGACCGCGATGAACAGGACAGCTCT 13469

RESULT 2
AV744150 15129 bp DNA circular SYN 03-DEC-2004
DEFINITION Dengue virus type 2 vector p4(delta30)-4995, complete sequence.
ACCESSION AV744150
VERSION AV744150.1 GI:56089728
KEYWORDS
SOURCE
ORGANISM
REFERENCE 1 (bases 1 to 15129)
AUTHORS Blaney, J.E. Jr., Hanson, C.T., Hanley, K.A., Murphy, B.R. and Whitehead, S.S.
TITLE Vaccine candidates derived from a novel infectious cDNA clone of an American genotype dengue virus type 2
JOURNAL BMC Infect. Dis. 4 (1), 39 (2004)
PUBMED 15461822
REFERENCE 2 (bases 1 to 15129)
AUTHORS Whitehead, S.S.
TITLE Direct Submission
JOURNAL Submitted (07-SEP-2004) Laboratory of Infectious Diseases, NIAID, 50 South Drive, Bldg 50, Room 6515, Bethesda, MD 20892, USA
FEATURES
source
1.15129
/organism="Dengue virus type 2 vector p4(delta30)-4995"
/mol_type="other DNA"
/strain="Tonga/74"
/db_xref="taxon:301537"
/country="Tonga"
/focus
1.10683
/organism="Dengue virus type 2"
/mol_type="other DNA"
/db_xref="taxon:11060"
10693.15054
/organism="Cloning vector pBR322"
/mol_type="other DNA"
/db_xref="taxon:47470"
97.10272
/codon_start=1
/product="polyprotein"
/protein_id="AAV70832.1"
```



```
Db 11942 TTGTTTCGGCGTGGGTATGGTGGCAGGCCCGCTGGCGGGGACATGTTGGGGCGCCATCTC 12001
Qy 1889 CTTGATGACCACTTCTTTCGGCGCGGTGTCTCAACGGCCCTCAACCTACTACTGCGCTG 1948
Db 12002 CTTGATGACCACTTCTTTCGGCGCGGTGTCTCAACGGCCCTCAACCTACTACTGCGCTG 12061
Qy 1949 CTTTCTTAATGACGAGTGCATTAAGGGAGAGCGTCGACCGATGCCCTTTGAGAGCCTTCAA 2008
Db 12062 CTTTCTTAATGACGAGTGCATTAAGGGAGAGCGTCGACCGATGCCCTTTGAGAGCCTTCAA 12151
Qy 2009 CCCAGTCAGCTCTTCTCCGCTGGCGCGGGGCATGACTATCGTCGCGGCACCTTATGACTGT 2068
Db 12122 CCCAGTCAGCTCTTCTCCGCTGGCGCGGGGCATGACTATCGTCGCGGCACCTTATGACTGT 12181
Qy 2069 CTTCTTTTATCATGCAACTCTGTATGAGACAGGTGCGGAGCGCTCTGGGTCAATTTTCGGCGA 2128
Db 12182 CTTCTTTTATCATGCAACTCTGTATGAGACAGGTGCGGAGCGCTCTGGGTCAATTTTCGGCGA 12241
Qy 2129 GGACCGCTTTCGCTCGAGCGGACCATGATCGGCGCTGTGCTTGGGTATTCGGAATCTT 2188
Db 12242 GGACCGCTTTCGCTCGAGCGGACCATGATCGGCGCTGTGCTTGGGTATTCGGAATCTT 12301
Qy 2189 GCACGCCCTCGCTCAAGCTTTCGCTCACTGGTCCCGCCACCAACCGTTTCGGCGAGAAGCA 2248
Db 12302 GCACGCCCTCGCTCAAGCTTTCGCTCACTGGTCCCGCCACCAACCGTTTCGGCGAGAAGCA 12361
Qy 2249 GGCCATTAATCGCGGCATGGCGGCGGACGCGCTGGGCTACGTTCTGCTGGGTTTCGGAC 2308
Db 12362 GGCCATTAATCGCGGCATGGCGGCGGACGCGCTGGGCTACGTTCTGCTGGGTTTCGGAC 12421
Qy 2309 GCGAGCTGGATGGCTTCCCAATTAATGATTTCTTCGCTTCGGCGGCATCGGATGCC 2368
Db 12422 GCGAGCTGGATGGCTTCCCAATTAATGATTTCTTCGCTTCGGCGGCATCGGATGCC 12481
Qy 2369 CGCGTTGAGGCGCATGCTGTCAGGCGAGTAGATCAGCAGCATCAGGAGACGTTTCAAGG 2428
Db 12482 CGCGTTGAGGCGCATGCTGTCAGGCGAGTAGATCAGCAGCATCAGGAGACGTTTCAAGG 12541
Qy 2429 ATCGCTCGCGCTTTTACCAGCCTAACTTCGATCATTTGGACCGCTGATCGTCACGCGCAT 2488
Db 12542 ATCGCTCGCGCTTTTACCAGCCTAACTTCGATCATTTGGACCGCTGATCGTCACGCGCAT 12601
Qy 2489 TTATGCCGCTCGGCGAGCACATGAACGGGTGCGATGAGATTGAGGCGCGCCCTTATA 2548
Db 12602 TTATGCCGCTCGGCGAGCACATGAACGGGTGCGATGAGATTGAGGCGCGCCCTTATA 12661
Qy 2549 CTTGTCTGCTTCCCGCGTTCGCTCGCGGTGCGATGGAGCGGGCCACTCGACTGAAT 2608
Db 12662 CTTGTCTGCTTCCCGCGTTCGCTCGCGGTGCGATGGAGCGGGCCACTCGACTGAAT 12721
Qy 2609 GGAAGCGCGGCGACCTCGCTAAACGGATTACCACTCCAAGAAATTGGAGCCAATCAATTC 2668
Db 12722 GGAAGCGCGGCGACCTCGCTAAACGGATTACCACTCCAAGAAATTGGAGCCAATCAATTC 12781
Qy 2669 TTGGGAGAACTGTGAATGCGCAACCAACCCCTTGGGAGAACATATCCATCGCTCGCC 2728
Db 12782 TTGGGAGAACTGTGAATGCGCAACCAACCCCTTGGGAGAACATATCCATCGCTCGCC 12841
Qy 2729 ATCTCCAGCAGCGCACGCGGCGCATCTCGGCGAGCGTTGGTCTCTGG----- 2776
Db 12842 ATCTCCAGCAGCGCACGCGGCGCATCTCGGCGAGCGTTGGTCTCTGG----- 2776
Qy 2777 ----- 2776
Db 12902 ATGATCGTCTCTGCTGTTGAGGACCGGCTAGGCTGGCGGGTTGCCCTTACTGGTTAG 12961
Qy 2777 ----- 2776
Db 12962 CAGAAATGAATCACCAGTACGCGAGCGAAGTGAAGCGACTGTGCTGCAAAACGTTCTCG 13021
Qy 2777 ----- 2776
Db 13022 ACCTGAGCAACAACATGAATGGTCTTTCGGTTTTCGTTTCGTAAGATCTCGAAACGCGG 13081
Qy 2777 ----- 2776
Db 13082 AAGTCAGCGCCCTGCACCACTTATGTTCCGGATCTGCATCGCAGGATGCTGCTGGCTACCC 13141
Qy 2777 -----GCTGGCATTTGACCTCGCTGAGTGAATTTTCTC 2805
Db 13142 TGTGGAACACCTACATCTGTATTACGAAGCGCTGGCATTTGACCTGAGTGAATTTTCTC 13201
Qy 2806 TGGTCCCGCGCATCATACCGCAGTTGTTTACCCCTCACAAAGTTTCCAGTAACCGGCA 2865
Db 13202 TGGTCCCGCGCATCATACCGCAGTTGTTTACCCCTCACAAAGTTTCCAGTAACCGGCA 13261
Qy 2866 TGTTCATCATCATGTAACCGGTATCGTGAGCATCTCTCTCGTTTCATCGGTATCATTACC 2925
Db 13262 TGTTCATCATCAGTAACCGGTATCGTGAGCATCTCTCTCGTTTCATCGGTATCATTACC 13321
Qy 2926 CCCATGAACAGAAATCCCGCTTACACGAGGAGCATCAGTGACCAACAGGAAAAACCGCC 2985
Db 13322 CCCATGAACAGAAATCCCGCTTACACGAGGAGCATCAGTGACCAACAGGAAAAACCGCC 13381
Qy 2986 CTTAACATGGCCCGCTTTTATCAGAACCCAGACATTAACGCTTCTCGAGAACTCAACGAG 3045
Db 13382 CTTAACATGGCCCGCTTTTATCAGAACCCAGACATTAACGCTTCTCGAGAACTCAACGAG 13441
Qy 3046 CTGACGCGGATGAACAGGCGAGCTCT 3073
Db 13442 CTGACGCGGATGAACAGGCGAGCATCT 13469
RESULT 3
AY656170 15145 bp DNA circular SYN 14-JAN-2005
LOCUS Dengue virus type 3 vector p3(delta30), complete sequence.
DEFINITION AY656170
ACCESSION AY656170
VERSION AY656170.1 GI:50345912
KEYWORDS
SOURCE
ORGANISM
Dengue virus type 3 vector p3(delta30)
Dengue virus type 3 vector p3(delta30)
other sequences; artificial sequences; vectors.
REFERENCE
1 (bases 1 to 15145)
AUTHORS Blaney, J.E. Jr., Hanson, C.T., Firestone, C.Y., Hanley, K.A.,
Murphy, B.R. and Whitehead, S.S.
Genetically modified, live attenuated dengue virus type 3 vaccine
candidates
Am. J. Trop. Med. Hyg. 71 (6), 811-821 (2004)
JOURNAL
PUBMED 15642976
REFERENCE 2 (bases 1 to 15145)
AUTHORS Whitehead, S.S.
TITLE Direct Submission
JOURNAL Submitted (17-JUN-2004) Laboratory of Infectious Diseases, NIAID,
50 South Drive, Bldg 50, Room 6515, Bethesda, MD 20892, USA
FEATURES
Location/Qualifiers
source
1..15145
/organism="Dengue virus type 3 vector p3(delta30)"
/mol_type="other DNA"
/db_xref="taxon:285000"
/focus
source
1..2345
/organism="Dengue virus type 3"
/mol_type="other DNA"
/db_xref="taxon:11069"
/feature="Sleman/78"
2369..10699
/organism="Dengue virus type 3"
/mol_type="other DNA"
/db_xref="taxon:11069"
/feature="Sleman/78"
10709..15070
/organism="Cloning vector pBR322"
/mol_type="other DNA"
/db_xref="taxon:47470"
15071..15145
source
```



```
Db 13098 AAGTCAGCGCCCTGCACCATATATGTCGGAATCGATCGCAGGATGCTGCTGGCTACCC 13157
QY 2777 -----CCTGCGATGACCTGAGTATTTTCTC 2805
Db 13158 TGTGAAACACCTACATCTGTATTAACGAAGCCTGCGATGACCTGAGTATTTTCTC 13217
QY 2806 TGGTCCCGCGCATCATACCGCCAGTGTGTTTACCTCTACACGCTTCCAGTAAACCGGCA 2865
Db 13218 TGGTCCCGCGCATCATACCGCCAGTGTGTTTACCTCTACACGCTTCCAGTAAACCGGCA 13277
QY 2866 TGTTCATCATCAGTAAACCGCATCTCGTACGATCCTCTCTCGTTTCATCGGTATCATTAACC 2925
Db 13278 TGTTCATCATCAGTAAACCGCATCTCGTACGATCCTCTCTCGTTTCATCGGTATCATTAACC 13337
QY 2926 CCATGAACAGAAATCCCGCTTACACGAGGATCAGTGACAAACAGGAAAAACCGCC 2985
Db 13338 CCATGAACAGAAATCCCGCTTACACGAGGATCAGTGACAAACAGGAAAAACCGCC 13397
QY 2986 CTTACATGCGCCGCTTTATCAGAGCCAGCATTAACGCTTCTCGAGAAACTCAACGAG 3045
Db 13398 CTTACATGCGCCGCTTTATCAGAGCCAGCATTAACGCTTCTCGAGAAACTCAACGAG 13457
QY 3046 CTGACGCGGATGAACAGCAGGCTCT 3073
Db 13458 CTGACGCGGATGAACAGCAGGATCT 13485

RESULT 4
AY744148
LOCUS AY744148 15159 bp DNA circular SYN 03-DEC-2004
DEFINITION Dengue virus type 2 vector p2, complete sequence.
ACCESSION AY744148
VERSION AY744148.1 GI:56089724
KEYWORDS
SOURCE
ORGANISM
REFERENCE 1 (bases 1 to 15159)
AUTHORS Blaney, J.E. Jr., Hanson, C.T., Hanley, K.A., Murphy, B.R. and Whitehead, S.S.
TITLE Vaccine candidates derived from a novel infectious cDNA clone of an American genotype dengue virus type 2
JOURNAL BMC Infect. Dis. 4 (1), 39 (2004)
PUBMED 15461822
REFERENCE 2 (bases 1 to 15159)
AUTHORS Whitehead, S.S.
TITLE Direct Submission
JOURNAL Submitted (07-SEP-2004) Laboratory of Infectious Diseases, NIAID, 50 South Drive, Bldg 50, Room 6515, Bethesda, MD 20892, USA
FEATURES
source
1..15159
/organism="Dengue virus type 2 vector p2"
/mol_type="other DNA"
/strain="Tonga/74"
/db_xref="taxon:301535"
/country="Tonga"
/focus
1..10713
/organism="Dengue virus type 2"
/mol_type="other DNA"
/db_xref="taxon:11060"
10723..15084
/organism="Cloning vector pBR322"
/mol_type="other DNA"
/db_xref="taxon:47470"
97..10272
/codon_start=1
/product="polyprotein"
/protein_id="AAV70830.1"
/db_xref="GI:56089725"
/translation="MNNQRKARNTPPNMLKRRNRVSTVOOLTFRSLGMLQGRGPL
KLFWALVAFRLPIPTTAGILKKWGTIKKSKAINLVRGFRKEIGRMNLINRRRTV
GMIIMLTPTVMAFLTRNGEPHIVSRQEKGLLFTKDKGTNNCTLMAMDGLGELCE
```

```
mat_peptide
mat_peptide
mat_peptide
mat_peptide
mat_peptide
mat_peptide
mat_peptide
mat_peptide
mat_peptide
mat_peptide
```

```
/product="anchored capsid protein (anchC)"
97..396
/product="viral capsid protein (virc)"
439..936
/product="membrane precursor protein (prM)"
712..936
/product="membrane protein (M)"
937..2421
/product="envelope protein (E)"
2422..3477
/product="NS1 protein"
3478..4131
/product="NS2A protein"
4132..4521
/product="NS2B protein"
4522..6375
```

```
DTITYKCPFLKQNEPEDIDCNCNSTSTWVTYGTCTTTTGEHRRKRSVALVPHVGMGLE
PRSTWMSSEGAWHAQRIETWILRHPOFTTMAAILAYTIGTTHPQRLVIFLLTAIA
PSMTMFCIGISNDFEVGSGSWDIIVLEHSGCVTTMAKQNTLFDLPDLFTEAKQPA
LTKYCIETAKLTNTTDSRPTQGEPTQEDKRFVCKHSGNDTGHGKVEKTTPOSSI
IVTCAMFTCKNMCKEIKVQENLEYTVVIIPHSGEHAGNDTGHGKVEKTTPOSSI
TEALGTGVTVMSCSPRTGLDFNEMVLQMEDKAWLVRQWFLDLPPLWPGATQOG
SNWIOKETLVTFKPHAKQDVVVLGSGEAMHTALTATGATEIQSSGNGLTGHUKCR
LRMDKLQJLWSPHACTCKFKIVKIEAETQHTIVIRVOYEGDGPCKIPFIMLEK
ETWRLGLTVPNIIVTEKDSVPNIEAEPFPGDSYIIIGVEPQLKLDWFKGSSIQMLF
RTWRAKRAIMAILGTADWDFSLGGVFTSIGKALHQVFAGIYGAAGFVSGVMTKMLIG
VIITWGNRSSTLSVSLVLVGLVGMVQADSCVSVKNKLEKCGSIIFFVTD
NVHTTEQYKQPSPSKLSAIOKAHEGICGTSVTRLENLMMKQITSELNHLVSE
NEVKLTITMGIDIKGIMQKRSRLRPOPTLRISWTKGAKMLSTELNGLTIDGPE
TAECPNTNRAWNSLEVEDYFGFTTNIWLRLKQDFCDKLSMAAIKONRAVHAD
MGYITESALNDTWIKIEKASFTVSKCHWPKSHLTWSNGVLESEMIIPNFAFGVQHN
NRPGYVOTAGPMLKLEMDPFCBGTVVVTECNGRPSLTIRTTASGKILGTWMLFC
RSTCLPLRYRGEDGCWYGMERIPLKEKEENLVSLVTAHGHOIDNFSILGLMALFL
BEMLRVTGTHAIIILVAVSVTLITGNMSFRDLGRVMVVGATNTDDIGMGTLYAL
LAFRVRPTFAAGLLRLKLTSELKMTTIGIVLLSQSSIPEITILELTALALGMMVLK
MVRNMEKYQLAVTMAILCVPNVILQNAWVSCTTILAVSVPLLLTSSQQAADMIP
LALITKGLNPTAIPLTTLRTNKRKSWPLNEAIVAGVMSILASLLKNDIPMTGPLY
AGLLTVCYLTITGRSADLELERATDVKWDQAEISGSPILSITISSEGSMSIKNEE
BOTLITLIRGLVLSGLFPVSIPITAAWYLVKWKQKQAGVMDVPPPPVGGKAELE
DGAYRIKQKGLYSGIAGYVYKGTFTTHMHTVKGAVLMKMKRIEISWADVKDLI
SYGGWKLGEWKEGSEVQVLALEPGKNPRAVQTPKPLFRNTGTGTGIVASLDFSGTS
GSPIVDKKGVGLYNGVVTGSAIYSAIAQTEKSIEDNPEIEDDIFRRLTITMDL
HPGAGTKRYLPFAIVREAIRKGLRTLILAPRVVAAMEEALRGLPIRYQPAIRAEH
TGREIVDMCHATFTMLLSPIRVNNYLIIMDEAHFTDPASIAARGVISTRVEMGEA
AGITPMTATPGSRDPPFQSNAPIMDEEREIPERSWNSGHEWTDPKGTVFVPSIKT
GNDIAACLRKNGKAVIQLSRKTPDSEYVKTITNDWDFVTTDISEMGANFKAERVIDP
RRCWPKVILTDGEERVILAGPMPVTHSSAARRGIRIGNRNENDQIYIMGEPLNDE
DCAHKEAKMLDINTPEGIIPSLFEPREKVDADIGEYRLGEARKTFVDLMRERG
LPVWLAYKVAEGINADRWCFDGTNRNIOLEENVEIWTKEGERKJLKPWLDGAR
IYSDPLAKKEPFAAGKSLTLNLTETMGLRPTFMTOKARDNLAVLTPAEAGGK
AYNHASLSELTETLTLTLATVTCGIFPLFMSGRGKMTLGMCCIIITASIIILWY
AQIQPHIASIILIEPLIIVLIPPEKQPTQDNLTYVILITVVAANWEMGP
LEKTKDGLIGNATQSPESNILDILRPASATLIYAVATTIPTMLRHSIENSVNV
SLTAANQATVLMGLGKGPLSKMDIGVPLLAIQCYQVNPITITLALLLVAHYAII
GPLQAKATREAAKAAAGINKNPTVDGITYDIDLPIDYDPKFKQLQGVMLLVLCVF
QIMNRTTWCALCEALTATGVPSTWEGNPGRFWNTTAVSWANIIFRGSYLAKGILF
SINKNTSTRGTCNIGETLGEKWSRLNALKSEFOIYKSGIOEVDRTLAKEGKFR
GETDHAVERGSAKLRWPFERNLVTPEGVVDLGGCGGWSYCGGLKNVRVKGKLT
GGPHEEPIPMSTYTGWNLVRLQSGVDVFPPEKCDTLILLDIGESSNPPTVEAGRTL
VLNLVENLNNNTQFCVKVLPNPSVIERMETLQRYGGALVRNPLSRNSTHEMYW
SNASGNISSVNMISRLINRFTMRHKATYEPDVLGSGSRNVTGIESETPNLDIIGK
RIEKIKOEHSTWYDQDHPYKWAYHGSYETKOGSSWVNGVWLLTKPDVVPWM
KLEKIGYILREVSKKEGAMTADTDTRITIEDKNEMITTHMAGEHKLAEIIF
LHKLGYILREVSKKEGAMTADTDTRITIEDKNEMITTHMAGEHKLAEIIF
IFKSIQHLTASEEIAVODLVRGRLSRMAISGDCVVKPLDDRPARALTALNDMG
KVRKDIOQWEPSPRGWNTQVPCFCHFHFLIMKDGRTLVVPCNQDELIGRARIISQ
AGWSLRETAGLKSIAQWLSMYFHRDLRLAANAICAVPSHWIPTSRTTWSIHASH
EMWTTDMLTVNNRVWILENPMEDKTPVSEEEIPLYLGRKEDQWCGSLIGLTSRATW
AKNIQTAINQVRSILGNEEYTDYMPMSMKFREREAEAGLVW"
```

| | | | | | |
|--|-------|--|-------|--|--|
| | | /product="NS3 protein" | | 34.0%; Score 1636.8; DB 11; Length 15159; | |
| mat_peptide | | 6376..6756 | | Best Local Similarity 86.4%; Pred. No. 0; | |
| mat_peptide | | /product="NS4A protein" | | Matches 1943; Conservative 0; Mismatches 22; Indels 283; Gaps 1; | |
| mat_peptide | | 6757..6825 | | | |
| mat_peptide | | /product="2K protein" | | | |
| mat_peptide | | 6826..7569 | | | |
| mat_peptide | | /product="NS4B protein" | | | |
| mat_peptide | | 7570..10269 | | | |
| misc_feature | | /product="NS5 protein" | | | |
| misc_feature | | 10714..10722 | | | |
| promoter | | /note="KpnI/AgeI linker" | | | |
| promoter | | 15141..15159 | | | |
| promoter | | /note="SP6" | | | |
| ORIGIN | | | | | |
| Query Match 34.0%; Score 1636.8; DB 11; Length 15159; | | | | | |
| Best Local Similarity 86.4%; Pred. No. 0; | | | | | |
| Matches 1943; Conservative 0; Mismatches 22; Indels 283; Gaps 1; | | | | | |
| Qy | 1109 | TGTGGAACACCTACATCTGTAATTAACGAAGCATTTATGAAGCATTTATCAGGCTTAATGTT | 1168 | | |
| Db | 11252 | TGTTGAATACTATATCTCTCTTTTCAATATTTAGAGCATTTATCAGGCTTAATGTT | 11311 | | |
| Qy | 1169 | CTCATGACGGATACATATTTGAATGTTATTTAGAAAAATAAACAATAAGGGTTCCGGCC | 1228 | | |
| Db | 11312 | CTCATGACGGATACATATTTGAATGTTATTTAGAAAAATAAACAATAAGGGTTCCGGCC | 11371 | | |
| Qy | 1229 | ACATTTCCCGAAAAAGTCCACTGACGCTTAAGAAACCATTTATTCATGACATTAAACC | 1288 | | |
| Db | 11372 | ACATTTCCCGAAAAAGTCCACTGACGCTTAAGAAACCATTTATTCATGACATTAAACC | 11431 | | |
| Qy | 1289 | TATAAAAATAGCGGTATCACGAGGCCCTTTGCTCTTCAAGAAATCTCATGTTTGACAGCT | 1348 | | |
| Db | 11432 | TATAAAAATAGCGGTATCACGAGGCCCTTTGCTCTTCAAGAAATCTCATGTTTGACAGCT | 11491 | | |
| Qy | 1349 | TATCATCGATAAGCTTTAATGCGGTAGTTTATCACAGTTTAAATTCGTAACGAGTCAGGC | 1408 | | |
| Db | 11492 | TATCATCGATAAGCTTTAATGCGGTAGTTTATCACAGTTTAAATTCGTAACGAGTCAGGC | 11551 | | |
| Qy | 1409 | ACCGTGTATGAATCTAAACATGCGCTCATCGTCACTCTCGGCACCGTCAACCTGGATGC | 1468 | | |
| Db | 11552 | ACCGTGTATGAATCTAAACATGCGCTCATCGTCACTCTCGGCACCGTCAACCTGGATGC | 11611 | | |
| Qy | 1469 | TGTAGGCATAGGCTTGGTTATGCGCGTACTGCGGCGCTCTTTCGGGATATCGTCCATTCT | 1528 | | |
| Db | 11612 | TGTAGGCATAGGCTTGGTTATGCGCGTACTGCGGCGCTCTTTCGGGATATCGTCCATTCT | 11671 | | |
| Qy | 1529 | GCACAGCATCGCCAGTCACTATGGCGTGTCTAGCGCTATATGCGTTTGATGCAATTTCT | 1588 | | |
| Db | 11672 | GCACAGCATCGCCAGTCACTATGGCGTGTCTAGCGCTATATGCGTTTGATGCAATTTCT | 11731 | | |
| Qy | 1589 | ATGCGCACCGCTTCTCGAGCACTGTCCGACCGCTTTGGCGCGCCAGTCTCTGCTGC | 1648 | | |
| Db | 11732 | ATGCGCACCGCTTCTCGAGCACTGTCCGACCGCTTTGGCGCGCCAGTCTCTGCTGC | 11791 | | |
| Qy | 1649 | TTGCGCTACTTGGAGCCACTATCGACTACGCGATCATGGCGACCAACCCGCTCTGTGGAT | 1708 | | |
| Db | 11792 | TTGCGCTACTTGGAGCCACTATCGACTACGCGATCATGGCGACCAACCCGCTCTGTGGAT | 11851 | | |
| Qy | 1709 | CCTCTAGCGCGGACGATCTGTGGCGGCGATCACCGGCGCCAGGTGGTGGCGC | 1768 | | |
| Db | 11852 | CCTCTAGCGCGGACGATCTGTGGCGGCGATCACCGGCGCCAGGTGGTGGCGC | 11911 | | |
| Qy | 1769 | CTATATCGCCGACATCACCGATGGGGAAGATCGGGCTCGCCACTTCGCGGCTCATGAGCG | 1828 | | |
| Db | 11912 | CTATATCGCCGACATCACCGATGGGGAAGATCGGGCTCGCCACTTCGCGGCTCATGAGCG | 11971 | | |
| Qy | 1829 | TTGTTTTCGGCGTGGGTATGGTGGCAGGCCCGTGGCGGGGACCTGTTGGGGCCATCTC | 1888 | | |
| Db | 11972 | TTGTTTTCGGCGTGGGTATGGTGGCAGGCCCGTGGCGGGGACCTGTTGGGGCCATCTC | 12031 | | |
| Qy | 1889 | CTTGATGCACTACCACTTCTTTCGGCGGCGGTGCTCAACGGCCCTCAACTACTACTGGGCTG | 1948 | | |

| | | | | | |
|----|-------|--|-------|--|--|
| Db | 12032 | CTTGCATGCACCAATTCCTTTCGCGCGCGGTCTCAACGGGCTCAACCTACTACTTGGGCTG | 12091 | | |
| Qy | 1949 | CTTCTAATGCAGAGTCGATAGGAGAGAGCGTCGACCGATGCTTGTAGAGAGCTTCAA | 2008 | | |
| Db | 12092 | CTTCTAATGCAGAGTCGATAGGAGAGAGCGTCGACCGATGCTTGTAGAGAGCTTCAA | 12151 | | |
| Qy | 2009 | CCAGTCAAGTCTCTTCGCGTGGGCGCGGGGATGACTATCGTTCGCGGACCTTATGACTGT | 2068 | | |
| Db | 12152 | CCAGTCAAGTCTCTTCGCGTGGGCGCGGGGATGACTATCGTTCGCGGACCTTATGACTGT | 12211 | | |
| Qy | 2069 | CTTCTTTATCATGCAACTCGTATGACAGGTGCGGAGCGCTCTGGGTCAATTTTCGGGA | 2128 | | |
| Db | 12212 | CTTCTTTATCATGCAACTCGTATGACAGGTGCGGAGCGCTCTGGGTCAATTTTCGGGA | 12271 | | |
| Qy | 2129 | GGACCGCTTTCGCTGGAGCGGACGATGATCGGCTGTGCTTCGGGTATTCGGAATCTT | 2188 | | |
| Db | 12272 | GGACCGCTTTCGCTGGAGCGGACGATGATCGGCTGTGCTTCGGGTATTCGGAATCTT | 12331 | | |
| Qy | 2189 | GCAGCGCTCTCGCTCAAGCTTTCGTCACCTGGTCCCGCCACCAACAGTTTCGCGGAGAAGCA | 2248 | | |
| Db | 12332 | GCAGCGCTCTCGCTCAAGCTTTCGTCACCTGGTCCCGCCACCAACAGTTTCGCGGAGAAGCA | 12391 | | |
| Qy | 2249 | GGCCATTATGCGCGGCAATGCGCGGACGCGCTGGGCTACGCTTGTCTGGGCTTCGCGAC | 2308 | | |
| Db | 12392 | GGCCATTATGCGCGGCAATGCGCGGACGCGCTGGGCTACGCTTGTCTGGGCTTCGCGAC | 12451 | | |
| Qy | 2309 | GGGAGGCTGGATGGCTTCCCATTTATGATTTCTTCGCTTCGCGGCGCATCGGATGCC | 2368 | | |
| Db | 12452 | GGGAGGCTGGATGGCTTCCCATTTATGATTTCTTCGCTTCGCGGCGCATCGGATGCC | 12511 | | |
| Qy | 2369 | CGCGTTGCAGGCCATGCTGTCAGGCGAGTAGATGAAGACCATCAGGAGACAGCTTCAAGG | 2428 | | |
| Db | 12512 | CGCGTTGCAGGCCATGCTGTCAGGCGAGTAGATGAAGACCATCAGGAGACAGCTTCAAGG | 12571 | | |
| Qy | 2429 | ATCGCTCGCGCTCTTACAGCCTAACTTCGATCAATTGGAACGCTGATCGTCAAGCGGAT | 2488 | | |
| Db | 12572 | ATCGCTCGCGCTCTTACAGCCTAACTTCGATCAATTGGAACGCTGATCGTCAAGCGGAT | 12631 | | |
| Qy | 2489 | TTATGCGGCTTCGGGAGCACATGGAACGGGTGCGATGATTTGAGGCGCGCCCTATA | 2548 | | |
| Db | 12632 | TTATGCGGCTTCGGGAGCACATGGAACGGGTGCGATGATTTGAGGCGCGCCCTATA | 12691 | | |
| Qy | 2549 | CTTGTCTGCGCTCCCGCGTTCGCGTGCATGGAACCGCGGCGCACCTCGACCTGAAT | 2608 | | |
| Db | 12692 | CTTGTCTGCGCTCCCGCGTTCGCGTGCATGGAACCGCGGCGCACCTCGACCTGAAT | 12751 | | |
| Qy | 2609 | GGAGCGCGGCGCACCTCGCTAAACGATTCACCATCCAGAAATGGAGGCAATCAATTC | 2668 | | |
| Db | 12752 | GGAGCGCGGCGCACCTCGCTAAACGATTCACCATCCAGAAATGGAGGCAATCAATTC | 12811 | | |
| Qy | 2669 | TTGCGGAGAACTGTGAAATGCGCAACCAACCTTGGGAGAACATATCCATCGCGTCCGCC | 2728 | | |
| Db | 12812 | TTGCGGAGAACTGTGAAATGCGCAACCAACCTTGGGAGAACATATCCATCGCGTCCGCC | 12871 | | |
| Qy | 2729 | ATCTCCAGCAGCGCACCGCGCGCATCTCGGCGAGCGTGGGTCTCTGG | 2776 | | |
| Db | 12872 | ATCTCCAGCAGCGCACCGCGCGCATCTCGGCGAGCGTGGGTCTCTGGC | 12931 | | |
| Qy | 2777 | ----- | 2776 | | |
| Db | 12932 | ATGATCGTGTCTCTGCTGTGAGGACCGCGCTAGGCTGGCGGGTTCCTTACTGCTTAG | 12991 | | |
| Qy | 2777 | ----- | 2776 | | |
| Db | 12992 | CAGAAATGAATCAACGATAGCGAGGAAAGTGGAACGAGTGTGCTGTCGAAAACGCTCGG | 13051 | | |
| Qy | 2777 | ----- | 2776 | | |
| Db | 13052 | ACCTGAGCAACAAATGAATGCTCTTCGTTTCGTTAAAGTCTGAAAACGCGG | 13111 | | |
| Qy | 2777 | ----- | 2776 | | |
| Db | 13112 | AAGTCAGCGCCCTGCACCAATTTATGTTCCGGATCTGCATCGCAGGATGCTGTGGTACCC | 13171 | | |

QY 2777 -----CCTGGCATTGACCTGAGTGATTTTCTC 2805
Db 13172 TGTGAAACACCTACATCTGTATTAAAGAGCGCTGGCATGACCTGAGTGATTTTCTC 13231
QY 2806 TGGTCCCGCGCATCATACGCCAGTTGTTTACCTCACAAGTTTCCAGTAACCGGGCA 2865
Db 13232 TGGTCCCGCGCATCATACGCCAGTTGTTTACCTCACAAGTTTCCAGTAACCGGGCA 13291
QY 2866 TGTTCATCATCAGTAACCGGATCGTAGCATCTCTCTCGTTTTCATCGGTATCATTTACC 2925
Db 13292 TGTTCATCATCAGTAACCGGATCGTAGCATCTCTCTCGTTTTCATCGGTATCATTTACC 13351
QY 2926 CCCATGAACAGAAATCCCTTTACAGGAGGATCAGTGACCAACAGAAAAACCGCC 2985
Db 13352 CCCATGAACAGAAATCCCTTTACAGGAGGATCAGTGACCAACAGAAAAACCGCC 13411
QY 2986 CTTAACATGGCCCGCTTTATCAGAAGCCAGACATTAACGCTTCTCGAGAACTCAACGAG 3045
Db 13412 CTTAACATGGCCCGCTTTATCAGAAGCCAGACATTAACGCTTCTCGAGAACTCAACGAG 13471
QY 3046 CTGACGCGGATGAACAGGCGAGCTCT 3073
Db 13472 CTGACGCGGATGAACAGGCGAGCTCT 13499

RESULT 5
AY656169
LOCUS AY656169 15176 bp DNA circular SVN 14-JAN-2005
DEFINITION Dengue virus type 3 vector p3, complete sequence.
ACCESSION AY656169
VERSION AY656169.1 GI:50345911
KEYWORDS
SOURCE
ORGANISM
Dengue virus type 3 vector p3
Dengue virus type 3 vector p3
Other sequences: artificial sequences; vectors.
REFERENCE
1 (bases 1 to 15176)
AUTHORS Blaney, J.B. Jr., Hanson, C.T., Firestone, C.Y., Hanley, K.A.,
Murphy, B.R. and Whitehead, S.S.
TITLE Genetically modified, live attenuated dengue virus type 3 vaccine
candidates
JOURNAL Am. J. Trop. Med. Hyg. 71 (6), 811-821 (2004)
PUBMED 15642976
REFERENCE 2 (bases 1 to 15176)
AUTHORS Whitehead, S.S.
TITLE Direct Submission
JOURNAL Submitted (17-JUN-2004) Laboratory of Infectious Diseases, NIAID,
50 South Drive, Bldg 50, Room 6515, Bethesda, MD 20892, USA
FEATURES
source
1. .15176
/organism="Dengue virus type 3 vector p3"
/mol_type="other DNA"
/db_xref="taxon:284999"
/focus
source
1. .2345
/organism="Dengue virus type 3"
/mol_type="other DNA"
/db_xref="taxon:11069"
/note="Sleman/78"
2369. .10730
/organism="Dengue virus type 3"
/mol_type="other DNA"
/db_xref="taxon:11069"
/note="Sleman/78"
10740. .15101
/organism="Cloning vector pBR322"
/mol_type="other DNA"
/db_xref="taxon:47470"
/note="Sleman/78"
source
15102. .15176
/organism="Cloning vector pGEM-3"
/mol_type="other DNA"
/db_xref="taxon:90108"
95. .10287
misc_feature

/note="encodes polypeptide precursor once SpeI linker
sequence is removed prior to preparation of infectious
transcripts corresponding to the viral genome"
2346. .2368
/note="SpeI linker; required for stable replication in
Escherichia coli"
10731. .10739
/note="KpnI/AgeI linker"
15158. .15176
/note="Sp6"

ORIGIN

Query Match 34.0%; Score 1636.8; DB 11; Length 15176;
Best Local Similarity 86.4%; Pred. No. 0;
Matches 1943; Conservative 0; Mismatches 22; Indels 283; Gaps 1;
QY 1109 TGTGAAACACCTACATCTGTATTAAAGAGCATTTATGAAGCATTTATCAGGTTATTGT 1168
Db 11269 TGTGAAATCTCATCTCTCTCTTTTCAATATTATTGAAGCATTTATCAGGTTATTGT 11328
QY 1169 CTGATGCGGATACATATTTGAATGTATTAGAAAAATAAACAATAGGGTTCCGGCG 1228
Db 11329 CTGATGCGGATACATATTTGAATGTATTAGAAAAATAAACAATAGGGTTCCGGCG 11388
QY 1229 ACATTTCCCGAAAGTCCACCTGACCTCTAAAGAACCATTTATTATCATGACATTAAAC 1288
Db 11389 ACATTTCCCGAAAGTCCACCTGACCTCTAAAGAACCATTTATTATCATGACATTAAAC 11448
QY 1289 TATTAATAGCGGTATCAGGAGCCCTTTTCGTCTTCAAGAAATTCATGTTTGACGCT 1348
Db 11449 TATTAATAGCGGTATCAGGAGCCCTTTTCGTCTTCAAGAAATTCATGTTTGACGCT 11508
QY 1349 TATCATCGATAAGCTTTTAATCGGTAGTTTATCACAGTTAAATTCCTAACGAGTCAGGC 1408
Db 11509 TATCATCGATAAGCTTTTAATCGGTAGTTTATCACAGTTAAATTCCTAACGAGTCAGGC 11568
QY 1409 ACCGTGTATGAAATCTAACAATGCGCTCATCGTCAATCTTCGCGACCGTCAACCTGGATGC 1468
Db 11569 ACCGTGTATGAAATCTAACAATGCGCTCATCGTCAATCTTCGCGACCGTCAACCTGGATGC 11628
QY 1469 TGTAGGATAGCGTTGGTTATGCCGGTACTGCCGGGCTCTTTCGGGATATGTCCTCAATTC 1528
Db 11629 TGTAGGATAGCGTTGGTTATGCCGGTACTGCCGGGCTCTTTCGGGATATGTCCTCAATTC 11688
QY 1529 CGACAGCATCGCCAGTCACTATGCGGTCTCTAGCGCTATATGCGTTGATGCAATTTCT 1588
Db 11589 CGACAGCATCGCCAGTCACTATGCGGTCTCTAGCGCTATATGCGTTGATGCAATTTCT 11748
QY 1589 ATGCGACCCGTTTCGAGACACTGTGCGACCGCTTTTGGCGCGCCAGTCTGCTCGC 1648
Db 11749 ATGCGACCCGTTTCGAGACACTGTGCGACCGCTTTTGGCGCGCCAGTCTGCTCGC 11808
QY 1649 TTGCTACTTGGAGCCACTATCGACTACCGCATATGCGGACCAACACCGCTCTGTGAT 1708
Db 11809 TTGCTACTTGGAGCCACTATCGACTACCGCATATGCGGACCAACACCGCTCTGTGAT 11868
QY 1709 CCTCTAGCGCGAGCATCTGTCGGCGCATCACCGGCGCACAGTGCGGTGCTGTCGCGC 1768
Db 11869 CCTCTAGCGCGAGCATCTGTCGGCGCATCACCGGCGCACAGTGCGGTGCTGTCGCGC 11928
QY 1769 CTATATCGCGACATCACCGATGGGGAAGATCGGGCTCGCACTTCGGGCTCATGAGCGC 1828
Db 11929 CTATATCGCGACATCACCGATGGGGAAGATCGGGCTCGCACTTCGGGCTCATGAGCGC 11988
QY 1829 TTGTTTCGGGTGGGTATGTTGGAGCGCCCGTGGCGGGGACTGTTGGGCGCCATCTC 1888
Db 11989 TTGTTTCGGGTGGGTATGTTGGAGCGCCCGTGGCGGGGACTGTTGGGCGCCATCTC 12048
QY 1889 CTTGCATGCACATTCCTTTCGGCGGGGCTCTCAACGGCTCTCAACCTACTACTTGGGCTG 1948
Db 12049 CTTGCATGCACATTCCTTTCGGCGGGGCTCTCAACGGCTCTCAACCTACTACTTGGGCTG 12108
QY 1949 CTTCTAATGAGGAGTCGCATAAGGGAGAGCGTCGACCGATGCCCTTGGAGGCTTCA 2008

```
Db 12109 CTTCTTAATGAGAGTCGATAGGAGAGCGTCGACCGATGCCCTTGAGAGCCTTCAA 12168
Qy 2009 CCAGTCAGCTCCTTCGCTGGCGCGCGGAGCA TGACTATGTCGCGCGCACTTATGACTGT 2068
Db 12169 CCCAGTCAGCTCCTTCGCTGGCGCGCGGAGCA TGACTATGTCGCGCGCACTTATGACTGT 12228
Qy 2069 CTTCTTTATCATGCAACTCGTAGACAGGTGCGCGGAGCGCTCTGGGTCATTTTCGGCGGA 2128
Db 12229 CTTCTTTATCATGCAACTCGTAGACAGGTGCGCGGAGCGCTCTGGGTCATTTTCGGCGGA 12288
Qy 2129 GGACCGCTTTCGCTGGAGCGGACGATGATCGGCGCTGTCTGCTCGGTCATTTTCGGAATCTT 2188
Db 12289 GGACCGCTTTCGCTGGAGCGGACGATGATCGGCGCTGTCTGCTCGGTCATTTTCGGAATCTT 12348
Qy 2189 GCACGCGCTCGCTCAAGCTTTCGCTACCTGTCGCGGACCAACGATGTCGGCGGAGGACA 2248
Db 12349 GCACGCGCTCGCTCAAGCTTTCGCTACCTGTCGCGGACCAACGATGTCGGCGGAGGACA 12408
Qy 2249 GGCCATTATCGCGGCGATGGCGGCGGACGCGCTGGGCTACGTCCTGCTGGCGTTTCGGGAC 2308
Db 12409 GGCCATTATCGCGGCGATGGCGGCGGACGCGCTGGGCTACGTCCTGCTGGCGTTTCGGGAC 12468
Qy 2309 GCGAGGCTGATGCGCTTCCCAATTATGATTTCTCTGCTTCCGCGCGGCAATCGGGATGCC 2368
Db 12469 GCGAGGCTGATGCGCTTCCCAATTATGATTTCTCTGCTTCCGCGCGGCAATCGGGATGCC 12528
Qy 2369 CGCGTTGAGGCCATGCTGTCAGGCGAGGTAGATGACGACCATCAGGACAGCTTCAAGG 2428
Db 12529 CGCGTTGAGGCCATGCTGTCAGGCGAGGTAGATGACGACCATCAGGACAGCTTCAAGG 12588
Qy 2429 ATCGCTCGCGCTTACCAGCCTTAATTCATCATTTGGACCGCTGATCGTCACGCGGAT 2488
Db 12589 ATCGCTCGCGCTTACCAGCCTTAATTCATCATTTGGACCGCTGATCGTCACGCGGAT 12648
Qy 2489 TTATGCGCCCTCGGAGCACATGAAACGGGTTGGCATGATTTAGCGCGCGCCCTATA 2548
Db 12649 TTATGCGCCCTCGGAGCACATGAAACGGGTTGGCATGATTTAGCGCGCGCCCTATA 12708
Qy 2549 CTTTGTCTGCTCCCGCTGTCGCTGCGGTGCAATGGAGCGCGGCGACCTCGACTGANT 2608
Db 12709 CTTTGTCTGCTCCCGCTGTCGCTGCGGTGCAATGGAGCGCGGCGACCTCGACTGANT 12768
Qy 2609 GGAACCGCGCGCACTTCCTAAACGGATTCACCACTCCCAAGAAATGGAGCCAATCAATTC 2668
Db 12769 GGAACCGCGCGCACTTCCTAAACGGATTCACCACTCCCAAGAAATGGAGCCAATCAATTC 12828
Qy 2669 TTGGGAGAACTGTGAATGCGCAAAACCAACCCCTTGGCAGAACATATCCATCGCGTCCGdC 2728
Db 12829 TTGGGAGAACTGTGAATGCGCAAAACCAACCCCTTGGCAGAACATATCCATCGCGTCCGdC 12888
Qy 2729 ATCTCCAGACGCGCACGCGCGCATCTCGGCGAGCGTTGGGTCCTGG----- 2776
Db 12889 ATCTCCAGACGCGCACGCGCGCATCTCGGCGAGCGTTGGGTCCTGGCCACCGGTCGdC 12948
Qy 2777 ----- 2776
Db 12949 ATGATCGTGTCTCTGTCTGATGAGACCGGCTAGGCTGCGGCGGTGCTTACTGTGTTAG 13008
Qy 2777 ----- 2776
Db 13009 CAGAAATGAATCACCGATACGCGAGCGAAGTGAAGGACTGCTGTGCAGAAACGTCCTGdG 13068
Qy 2777 ----- 2776
Db 13069 ACCTGAGCAACAAATGAATGCTCTCGGTTTCCGTTTTCGTAAGTCTGGAACCGCG 13128
Qy 2777 ----- 2776
Db 13129 AAGTCAGCGCCCTGCACCAATTATGTCGGATCTGCATCGCAGGATGCTGTCGGCTACCC 13188
Qy 2777 -----GCTGGCATTCACCTGAGTGAATTTTCTC 2805
|||||
```

```
Db 13189 TGTGAACACCTACATCTGTATTAAAGAACGCTGGCATTGAGCCCTGAGTGAATTTTCTC 13248
Qy 2806 TGTGCCGCGCATCATACCGCAGTTGTTTACCCCTCACAAGTTCACAGTAAACCGGCA 2865
Db 13249 TGTGCCGCGCATCATACCGCAGTTGTTTACCCCTCACAAGTTCACAGTAAACCGGCA 13308
Qy 2866 TGTTCATCATCAGTAACCCCGTATCGTGAGCATCTCTCTCGTTTCATCGGTATCATTACC 2925
Db 13309 TGTTCATCATCAGTAACCCCGTATCGTGAGCATCTCTCTCGTTTCATCGGTATCATTACC 13368
Qy 2926 CCATGAACAGAAATCCCTTTACACGGAGGCATCAGTACCAACAGGAAAAACCGCC 2985
Db 13369 CCATGAACAGAAATCCCTTTACACGGAGGCATCAGTACCAACAGGAAAAACCGCC 13428
Qy 2986 CTTAAACATGCCCGCTTTATCAGAAAGCAGACATTAACGCTTCTGGAGAACTCAACGAG 3045
Db 13429 CTTAAACATGCCCGCTTTATCAGAAAGCAGACATTAACGCTTCTGGAGAACTCAACGAG 13488
Qy 3046 CTGACCGCGATGAACAGGCGAGCTCT 3073
Db 13489 CTGACCGCGATGAACAGGCGAGCATCT 13516
```

```
RESULT 6
AY243469
LOCUS AY243469 15237 bp DNA circular SYN 20-OCT-2003
DEFINITION Chimeric Dengue virus vector p4 (Delta30) -D2-CME, complete sequence.
ACCESSION AY243469
VERSION AY243469.1 GI:30026605
KEYWORDS
SOURCE
ORGANISM Chimeric Dengue virus vector p4 (Delta30) -D2-CME
REFERENCE Chimeric Dengue virus vector p4 (Delta30) -D2-CME
AUTHORS other sequences; artificial sequences; vectors.
TITLE
JOURNAL Whitehead, S.S., Hanley, K.A., Blaney, J.E. Jr., Gilmore, L.E.,
PUBMED Substitution of the structural genes of dengue virus type 4 with
REFERENCE those of type 2 results in chimeric vaccine candidates which are
JOURNAL attenuated for mosquitoes, mice, and rhesus monkeys
Vaccine 21 (27-28), 4307-4316 (2003)
14505913
2 (bases 1 to 15237)
Whitehead, S.S.
Direct Submission
Submitted (24-FEB-2003) LID, NIAID, 50 South Dr., Room 6515,
Bethesda, MD 20892, USA
Location/Qualifiers
FEATURES
source
1. .15237
/organism="Chimeric Dengue virus vector
p4 (Delta30) -D2-CME"
/mol_type="genomic DNA"
/db_xref="taxon:227158"
/focus
1. .88
/organism="Dengue virus type 4"
/mol_type="genomic DNA"
/db_xref="taxon:11070"
89. .2343
/organism="Dengue virus type 2 (NGC-prototype)"
/mol_type="genomic DNA"
/db_xref="taxon:11065"
2344. .10616
/organism="Dengue virus type 4"
/mol_type="genomic DNA"
/db_xref="taxon:11070"
97. .10263
/codon_start=1
/product="polyprotein precursor"
/protein_id="AAP06255.1"
/db_xref="GI:30026606"
/translation="MNNORKKARNTPFNLKRNERNVSTVOQLTKRFSIGMLQGRGPL
KLFWALVAFRLPTTPAGILKRWGTLTKSKALNVLKRGFRKIGRMNLNRRRTA
GMIIMLPTVMAFHLITRNGEPHIVSRQEKGSLLFKTEGDNVNMCTLWAMDIGELCE
```


DTITVYKCLLRNEPDDCKCNSTSTWVYCTCTCTTCEHREKRSVALVPHVGMGLE
TRTWTMSSSEGAWKHAQRIETWILRHGFTIMAAIATYITGTHFORALIFILITAVA
PSMTWRCIGISNRDVEVSGSWVDIVLEHSGCVTWTAKNPLTDLFELITEAKQAP
TLRKYCIFAKLNTTTSRPTQESPLNEBQDKRFVCKHSMVDRGNGGCLGFKGG
IVTCAMFTCKNMEGVQVQENLEYITI VTPHSEBAHVNDGTGKHKEIKITPOSSI
TEAELTGVTGTFKPHAKQDVVVLGSGOEGAMHTALCATSEIQMSGNLLITGHLKCR
SNMLQKETLVTIPKPHAKQDVVVLGSGOEGAMHTALCATSEIQMSGNLLITGHLKCR
LMDKLQKHSYSCMCTCKFKVKEIAETQHGTVIRVOYEGDGSFKPIPEIMDLK
RVLGRLITVNPITVEKDSVNIIEABPPFGDSYIIIGVEPQLKLNFKFGSITGMF
ETWRGAKMADFDWFGSLGCVFTSIGKALHQVFGAIFYGAFFSGVSWTMKILIG
VITWIGNSRNTSAMTCIAVGGITLFLFTVQADMGCVASGKELCKGSGIFVVD
NVHTWEOYKQPPESPARLASAILNAHDKGCVGIRSTRLENVVMKQITNENLVLM
GSHDLTVVAGDVKGVLTGKALTPPVSDLYSKWTGKAKIFTPPEARNSFLIDGPD
TSECPNERANNSLEVEDYFGFMFTINIMKPFSGSEVCDHRLHSAAKQKAVHAD
MGYWIESKNWQIWEKASLIEVTKTLMPKTHTLWSNGVLESQMLIPKSYAGFPFQHN
RSCYGTQVTPHGLKLEIDFEGCPGTVTIQEDCDHRGSLRRTTASGKLVTQWCC
RSTWMPPLRFEGDCGWCMEIRPLSEKEENNVKSQVTAQGTSETSMGLCLTFLV
BECLRVRTRKMLILVVITLCAIILGGLTMDLLRALIMLGDMSGRIGQIHLAM
AVFKMSPGVILGVFLKLTSTETALMVGMAITVLSIPDLMLIDIGISGLILIKI
VTQDNTQVGTALSLTPIRSTMLVMAWRTIMAVFVVTLLPICTSCLOKQSHWE
ITALILGAQALPVYMLTMKASRRSWPLNEGIMAVGLSLLGSALLKNDVPDLAGPV
AGGLLAAVYMSGSSADLSLEKAAVQMDENADITGSSPIIEVKQDEDSFIRDVBE
TNMTLTLVLALEITVSGLYPLAIPVTMTLWYMWQVKTORSALMDVPSPAATKAALS
EGVYRIMORGLFGKTOGVGTHMEGVFHTMHVTRGVSICHETGRLEPSADVENDMI
SYGGMRGLDKWDEEVQVLAIEPGKNPKHVQTKPGLFKTLTSEIGNVTLDFPGTIS
GSPILNRKGVILGYNRGLVTKSGDYSAITQAEIRIGEPEYDEVEDIFRKRRLTIMDL
HPGAGTKRILPLSVIRALKRRLTLTILAPTRVVAAMEEALRGLPIRYQTPVAKSE
GREIIVDMCHATEFTTLSSRVPNTVNLIPEREAHFTDPSVAARGVISTRVPMSEA
AAI FTWATPPCATDPPFQNSPIEDIEREIPERSWNTGFDWITDYQGTVMFVPSIKA
GNDIANCLRGSKGVIOISRTKPTFTEPKTLTDWDFVVTDDISEMGANFRAGRVIDP
RCLKPLVLPDGPVILAGPIVTPASAAQRGRIGRNPAQEDDQVVFSGDPLANDE
DHAWTEAKMLDNIITPEGIIPILFGPEREKTOAIDGFEPLRGEORRTFVELMRRGD
LPWLSYKVASAGISYKDRWCFTGRNNOILEENMEVEIWTEGEEKKLRPWLDRAR
VYADPMALKDPFASGRKSTLIDILTEIASLPTLSSRAKLALDNI VMLHTTBERGR
AYOHALNELPSLETLMLVALLGAMTAGIPLFFMOGIGIKLSMGLITIAVASGLLW
AIEIOPWIAASIILIEFLMVLIIPEPEKQTPQDNOQLIYVLTILITIGLIAANMGL
IEKTKTDFGFQVKTETILDVLRPASAWTLYAVATILTPMERHTIENTSALSLSA
LANQAVALMGLGKWPVLRHMDLGVPLIAMGCVSQVNPPTLTASLMLLVHYAIIIGP
LOAKATREAQRTAAGIKMKNPTVDGIVTIDLEPISYDPKFKQLGVMLLVLCAQOLL
LMRTTFAFCEVTLATGPIILTLWEGNPRFWNTTIAVSTANI PRGSLAGAGLAFSLI
KNATPRRGCTTCTETLGEKWKROLNSLDKREFEYKRSGLILEVDRTEAKSALDKGSK
IKHVASGSSKIRMI VERGVMPKGVVDLGCGRGWSYMATLKNTEVKGTYTKGP
GHEPIIPMATYGNLVLKHSVDVIFYKTEQVDTLLCDIGESSNPTIEBGTURVLK
MVEPLSSKPEFCIKVLNPMYPTIELEKILQRKHGNNVRCPLSRNSTHEMYVWSA
SGNTSVSNTTSMKLARFTTRHRKPTYEKDVLDLAGTRSVSTETEKPDMTIIGRLQ
RLQEHKEHTHYDQENPVRTWAYHGSYEAPOSTGSASWNGVVKLLTKPMDVI PMVTO
LAMTDTTPGQORVEKEVDTRTPQPKGTBMVNTTANMLWALGKKKPKLCTREE
FISKVRNALGAFQEQWTSASEAVNDSRFVELVXERLALHQBCKSCSVYNNMG
KREKKGFEKAGSRATWYMLGARLEFEALGDLNEDHWFGRNSWSVGEGLHR
LGYLLEBIDKQGLMADDTAGWDTRI TEDDLQNEELITEQMAPHKKILAKI FKLIT
YONKVKVLRTPRGA VMDIISRKDRGSGQVGTGYLNTFTNMEVQILIRQWAEVGT
QDDMNPKGLKERVGLKECVDRLKRMAISGDDCVVPLDERFGTSLLP LNDMGV
RKDIPOWPSKGNWQVPCSRHPHKIEMKGRSLVVPVCRNODLIGRAFLSQGAV
KSLBETACLKAYAQMSLMYFHRDRDLASMAICSAVPTPEWFTSRTTWSIHAHQW
MHTBDMUKVNRNIBDNPNMTDKTIPVHSMEIDIPYLGKREDLMCGSLIGLSSRATWAK
NIHTAITQVNRILGKEVYDMPVMKRYAPSESEGLV"

| | | | | | |
|-----------------------|-----------------|---|-------------|---|-------|
| mat_peptide | 3478..4131 | /product="NS2A protein" | 1109 | TGTGGAACACCTACATCTGTATTAAACGAAGCAATTTTGAAGCAATTTATCAGGGTTATTGT | 1168 |
| mat_peptide | 4132..4521 | /product="NS2B protein" | 1155 | TGTTGAATCACTACTCTCTCTTTTCAATATTATTGAAGCAATTTATCAGGGTTATTGT | 11214 |
| mat_peptide | 4522..6375 | /product="NS3 protein" | 1169 | CTCATGAGCGATACATATTTGAATGTATTAGAAAAATAAACAATAGGGGTTCCGGCG | 1228 |
| mat_peptide | 6376..6756 | /product="NS4A protein" | 11215 | CTCATGAGCGATACATATTTGAATGTATTAGAAAAATAAACAATAGGGGTTCCGGCG | 11274 |
| mat_peptide | 6757..6825 | /product="2K protein" | 1229 | ACATTTCCCCGAAAGTGCCACCTGAGCTCTAAAGAAACCAATTTATCATGACATTAAACC | 1288 |
| mat_peptide | 6826..7560 | /product="NS4B protein" | 11275 | ACATTTCCCCGAAAGTGCCACCTGAGCTCTAAAGAAACCAATTTATCATGACATTAAACC | 11334 |
| mat_peptide | 7561..110260 | /product="NS5 protein" | 1289 | TATAAAATAGCGCTATCAACGAGCGCCCTTTCGTCTTCAAGAAATTCCTCATGTTTGACAGCT | 1348 |
| ORIGIN | | | 11335 | TATAAAATAGCGCTATCAACGAGCGCCCTTTCGTCTTCAAGAAATTCCTCATGTTTGACAGCT | 11394 |
| Query Match | 34.0%; | Score 1636.8; | DB 11; | Length 15237; | |
| Best Local Similarity | 86.4%; | Pred. No. 0; | | | |
| Matches 1943; | Conservative 0; | Mismatches 22; | Indels 283; | Gaps 1; | |
| Qy | 1109 | TGTGGAACACCTACATCTGTATTAAACGAAGCAATTTTGAAGCAATTTATCAGGGTTATTGT | 1168 | | |
| Db | 1155 | TGTTGAATCACTACTCTCTCTTTTCAATATTATTGAAGCAATTTATCAGGGTTATTGT | 11214 | | |
| Qy | 1169 | CTCATGAGCGATACATATTTGAATGTATTAGAAAAATAAACAATAGGGGTTCCGGCG | 1228 | | |
| Db | 11215 | CTCATGAGCGATACATATTTGAATGTATTAGAAAAATAAACAATAGGGGTTCCGGCG | 11274 | | |
| Qy | 1229 | ACATTTCCCCGAAAGTGCCACCTGAGCTCTAAAGAAACCAATTTATCATGACATTAAACC | 1288 | | |
| Db | 11275 | ACATTTCCCCGAAAGTGCCACCTGAGCTCTAAAGAAACCAATTTATCATGACATTAAACC | 11334 | | |
| Qy | 1289 | TATAAAATAGCGCTATCAACGAGCGCCCTTTCGTCTTCAAGAAATTCCTCATGTTTGACAGCT | 1348 | | |
| Db | 11335 | TATAAAATAGCGCTATCAACGAGCGCCCTTTCGTCTTCAAGAAATTCCTCATGTTTGACAGCT | 11394 | | |
| Qy | 1349 | TATCATGATGAAGCTTTAATCGCGTAGTATTATCATCAGTTAAATTCGTAACGAGTCAGGC | 1408 | | |
| Db | 11395 | TATCATGATGAAGCTTTAATCGCGTAGTATTATCATCAGTTAAATTCGTAACGAGTCAGGC | 11454 | | |
| Qy | 1409 | ACCGTGATGAATCTAACATGCGCTCATCGTCTCCGCGCCGTCACCCCTGGATGC | 1468 | | |
| Db | 11455 | ACCGTGATGAATCTAACATGCGCTCATCGTCTCCGCGCCGTCACCCCTGGATGC | 11514 | | |
| Qy | 1469 | TGTAGGATAGGCTTGGTTATGCGCGTACTGCCGGGCTCTTGGGGGATATCGTCAATTC | 1528 | | |
| Db | 11515 | TGTAGGATAGGCTTGGTTATGCGCGTACTGCCGGGCTCTTGGGGGATATCGTCAATTC | 11574 | | |
| Qy | 1529 | CGACAGCATGCCAGTCACATATGCGCGTGTCTAGCGCTATATGCGTTGATGCAATTTCT | 1588 | | |
| Db | 11575 | CGACAGCATGCCAGTCACATATGCGCGTGTCTAGCGCTATATGCGTTGATGCAATTTCT | 11634 | | |
| Qy | 1589 | ATGCGCACCGCTTCTCGGAGCAGTCTCGGCGCTTTGGCGCGCCGCGCTCTGCTCGC | 1648 | | |
| Db | 11635 | ATGCGCACCGCTTCTCGGAGCAGTCTCGGCGCTTTGGCGCGCCGCGCTCTGCTCGC | 11694 | | |
| Qy | 1649 | TTCCGCTACTTGGAGCCACTATCGACTACGGCATCATGCGCGACCAACACCCGCTCTGTGGAT | 1708 | | |
| Db | 11695 | TTCCGCTACTTGGAGCCACTATCGACTACGGCATCATGCGCGACCAACACCCGCTCTGTGGAT | 11754 | | |
| Qy | 1709 | CCTTACCGCGGACGATCTGTGGCGGATACACCGCGCCACAGGTGCGTGTGCGCGC | 1768 | | |
| Db | 11755 | CCTTACCGCGGACGATCTGTGGCGGATACACCGCGCCACAGGTGCGTGTGCGCGC | 11814 | | |
| Qy | 1769 | CTATATCGCGACATCACCGATGGGGAAGATCGGGCTCGGCACCTCGGGCTCATGAGCGC | 1828 | | |
| Db | 11815 | CTATATCGCGACATCACCGATGGGGAAGATCGGGCTCGGCACCTCGGGCTCATGAGCGC | 11874 | | |
| Qy | 1829 | TTGTTTCGCGTGGGTATGTTGGCAGCGCCGCTGGCGGGGAGATGTTGGCGCGCATCTC | 1888 | | |
| Db | 11875 | TTGTTTCGCGTGGGTATGTTGGCAGCGCCGCTGGCGGGGAGATGTTGGCGCGCATCTC | 11934 | | |
| Qy | 1889 | CTTGATGCACCAATTCCTTTCGCGCGCGGTGCTCAACGGCCCTCAACCTACTACTGCGGTG | 1948 | | |

| | | | |
|------------|-------|---|-------|
| Db | 11935 | CTTGATGACCAATCTCTTGGGCGGCGTGTCTCAACGGGCTCAACCTACTACTGGGCTG | 11994 |
| Qy | 1949 | CTTCCTAATGAGAGTGCATAAAGGAGAGCGTGCACCGATGCCCTTGAGAGCCTTCAA | 2008 |
| Db | 11995 | CTTCCTAATGAGAGTGCATAAAGGAGAGCGTGCACCGATGCCCTTGAGAGCCTTCAA | 12054 |
| Qy | 2009 | CCGAGTCAGCTCCTTCCGGTGGGCGCGGGGATGACTATCGTGCAGCCACTTATGACTGT | 2068 |
| Db | 12055 | CCGAGTCAGCTCCTTCCGGTGGGCGCGGGGATGACTATCGTGCAGCCACTTATGACTGT | 12114 |
| Qy | 2069 | CTTCCTTATCATGCAACTCGTAGACAGAGTGCAGCGAGCGCTCTGGGTCACTTTTCGGCGA | 2128 |
| Db | 12115 | CTTCCTTATCATGCAACTCGTAGACAGAGTGCAGCGAGCGCTCTGGGTCACTTTTCGGCGA | 12174 |
| Qy | 2129 | GGACCGCTTTCCGCTGGAGCGGACGATGATCGGCTGTGCTTCGGTATTTGGAATCTT | 2188 |
| Db | 12175 | GGACCGCTTTCCGCTGGAGCGGACGATGATCGGCTGTGCTTCGGTATTTGGAATCTT | 12234 |
| Qy | 2189 | GCAAGCCCTCGCTCAAGCCTTCGTCACTGTCCTCCGCGACCAAAAGTTTCGGCGAGAAGCA | 2248 |
| Db | 12235 | GCAAGCCCTCGCTCAAGCCTTCGTCACTGTCCTCCGCGACCAAAAGTTTCGGCGAGAAGCA | 12294 |
| Qy | 2249 | GGCCATTATCGCGGCGATGGCGGCGGACGCGCTGGGCTACGCTTGTCTGGCGTTTCGGAC | 2308 |
| Db | 12295 | GGCCATTATCGCGGCGATGGCGGCGGACGCGCTGGGCTACGCTTGTCTGGCGTTTCGGAC | 12354 |
| Qy | 2309 | CGGAGGCTGGATGGCTTCCCATATATGATTTCTTCGCTTCCGCGGCGATCGGATGCC | 2368 |
| Db | 12355 | CGGAGGCTGGATGGCTTCCCATATATGATTTCTTCGCTTCCGCGGCGATCGGATGCC | 12414 |
| Qy | 2369 | CGGTTGAGCGCATGCTGTCCAGGAGGTAGATGACGACCATCAGGACAGCTTCAAGG | 2428 |
| Db | 12415 | CGGTTGAGCGCATGCTGTCCAGGAGGTAGATGACGACCATCAGGACAGCTTCAAGG | 12474 |
| Qy | 2429 | ATCGCTCGCGCTCTTACAGCCTAACTTCGATCATTCGACCGCTGATCGTCAGCGCAT | 2488 |
| Db | 12475 | ATCGCTCGCGCTCTTACAGCCTAACTTCGATCATTCGACCGCTGATCGTCAGCGCAT | 12534 |
| Qy | 2489 | TTATGCCGCTCCGGGACACATGAAACGGGTGTGCATGATGTGAGCGCGCCCTATA | 2548 |
| Db | 12535 | TTATGCCGCTCCGGGACACATGAAACGGGTGTGCATGATGTGAGCGCGCCCTATA | 12594 |
| Qy | 2549 | CCTGTCTGCTCCCGGTTGCTCGCGGTGTCATGGAGCGCGGACCTCGACTGAT | 2608 |
| Db | 12595 | CCTGTCTGCTCCCGGTTGCTCGCGGTGTCATGGAGCGCGGACCTCGACTGAT | 12654 |
| Qy | 2609 | GGAGCGCGGCGACCTCGCTAACGGATTCACCACTCCAAGATTTGGAGCCCAATCAATTC | 2668 |
| Db | 12655 | GGAGCGCGGCGACCTCGCTAACGGATTCACCACTCCAAGATTTGGAGCCCAATCAATTC | 12714 |
| Qy | 2669 | TTGCGGAGAACTGTGAATGCGCAACCAACCCCTTGGCAGAACATATCCATCGGTCGCGC | 2728 |
| Db | 12715 | TTGCGGAGAACTGTGAATGCGCAACCAACCCCTTGGCAGAACATATCCATCGGTCGCGC | 12774 |
| Qy | 2729 | ATCTCAGAGCGCGACGGGGGATCTCGGGGACGCTTGGTCTCTGG----- | 2776 |
| Db | 12775 | ATCTCAGAGCGCGACGGGGGATCTCGGGGACGCTTGGTCTCTGGGGTGGCGC | 12834 |
| Qy | 2777 | ----- | 2776 |
| Db | 12835 | ATGATCGTCTCTGTCTGTGAGGACCCGCGTAGGCTGGCGGGTGGCTTACTGCTTAG | 12894 |
| Qy | 2777 | ----- | 2776 |
| Db | 12895 | CAGATGAATCACCAGATACGCGAGCGAAAGTGAAGCGACTGCTGCTGCAGAAACGTCGCG | 12954 |
| Qy | 2777 | ----- | 2776 |
| Db | 12955 | ACCTGAGCAACAAATGATGATGCTTTCGGTTCCTGTTTCGTAAGTCTGGAAACGCGG | 13014 |
| Qy | 2777 | ----- | 2776 |
| Db | 13015 | AACTCAGCGCCCTGCACCACTATTGTTCCGGATCTGCATCGCAGATCTGCTGGCTACCC | 13074 |
| Qy | 2777 | -----GTTGGCATTGACCTGAGTGATTTTCTC | 2805 |
| Db | 13075 | TGTGGAACACTATACCTGTAATAACGAAGCGCTGGCAATTGACCTGAGTGATTTTCTC | 13134 |
| Qy | 2806 | TGTTCCCGCGCATCCATACCGCCAGTTGTTTACCCCTCACAACGTTCCAGTTAACCGGCA | 2865 |
| Db | 13135 | TGTTCCCGCGCATCCATACCGCCAGTTGTTTACCCCTCACAACGTTCCAGTTAACCGGCA | 13194 |
| Qy | 2866 | TGTTTCATCATCAGTAACCCGATATCGTGAGCATCTCTCTGTTTCATCGGTATCATTACC | 2925 |
| Db | 13195 | TGTTTCATCATCAGTAACCCGATATCGTGAGCATCTCTCTGTTTCATCGGTATCATTACC | 13254 |
| Qy | 2926 | CCCATGACAGAAATCCCTTACACGAGGCGATCAGTGACCAACAGGAAAAACCGCC | 2985 |
| Db | 13255 | CCCATGACAGAAATCCCTTACACGAGGCGATCAGTGACCAACAGGAAAAACCGCC | 13314 |
| Qy | 2986 | CTTAACATGCGCCGCTTTATCAGAAAGCAGACATTAACGCTTCTGGAGAACTCAACGAG | 3045 |
| Db | 13315 | CTTAACATGCGCCGCTTTATCAGAAAGCAGACATTAACGCTTCTGGAGAACTCAACGAG | 13374 |
| Qy | 3046 | CTGACCGCGATGAACAGGCGAGCTCT | 3073 |
| Db | 13375 | CTGACCGCGATGAACAGGCGAGCATCT | 13402 |
| RESULT 7 | | | |
| AY243467 | | | |
| LOCUS | | | |
| DEFINITION | | | |
| ACCESSION | | | |
| VERSION | | | |
| KEYWORDS | | | |
| SOURCE | | | |
| ORGANISM | | | |
| REFERENCE | | | |
| AUTHORS | | | |
| TITLE | | | |
| JOURNAL | | | |
| PUBMED | | | |
| REFERENCE | | | |
| AUTHORS | | | |
| TITLE | | | |
| JOURNAL | | | |
| FEATURES | | | |
| source | | | |
| source | | | |
| source | | | |
| source | | | |
| CDS | | | |

```
/translation="NQORKKVVRPPFNWMLKRRNRVSTPQGLVKRSTGLPSGKPLR
MVLAFITFLRVLSPTTAGIDLRWGQKLNKAIKILIGFRKEIGRMNLNLRKRSAG
MIIMLITVMAFHLITVDEPHMIVSRQEGKSLFKETEDGVNNCTLMAMDGLCELD
TITYKLLRNEPESDIDWCNSTRSTVWTYGTCTTGEHREKRSVALVRPHVNGLET
RTETWMSSEGAKHQAQETELWLRHPGFTTMAAILAVTIGTFHQRALIFLLTAVAP
SMTWRCIGISNRDVEFGVSGSWWDIVLEHSGSVTTWAKNKTLDLFIKTEAKQAT
LKYCIEAKLTNTTRESRCPQGSPLNEEQDKRPVCKHSMVDKNGNGKGFQKGGI
VTCANFTCKMNEKGVOPENLEYITIVTHSGBEHAVGNDTGRHGEIKITPOSSIT
EALITGTVTMECSPTGLDFNEMVLLQMKENKAWLVHRQWFLDLPLPFLPGADTQS
NIQKETLVTFNPHAKQDVVVLGSEGAHMTALTATGATEIQMSGNLLFTGHLKCLR
HMDKLQKNGSKTGKFKVVKIEAETQHGTVIRVYEGDGGSPCKIPKPIIMDLKLR
RVGLRLITVNPISVTEKDSPVNI EAEPPGDSYI IIGVEPGOLKNWPKKSSIGOMPE
TTRGAKMAILGTDANDPGLSGVFTS I GKALHGVFGAIYGAAPSGVSWTKMLIGV
IITWIGNSRNTSWMTACIAVGGITLFLGTVOADMGCVASWSGKELKCGSGIFPVND
VHTTTEQYKQPESPARLASALINAHKDGVCGRISITRLENVMWKQITNELNYLWEG
GHDLTVVAGDVKGLTGRKRALTPVSDLKYSWTKGWKAKI FTPEARNSTPLIDPOT
SECNERRAWSLVEVDYGFQMTNTIWMKPRFGSSVCDHRLMSAAI KQOKA VHAD
GYWLESSKNQWQIEKASLIEVKTLAPKQHTLMSNGVLBSQMLI PKSVAGPFSQHY
ROGATQTVGPHWLGKLEIDRGECPGTVTIQEDCDHRGFSRLTITASGKLVQWCCR
SCTMPPFLFGECDGSGMEIRPLSEKEENNVKSQVTAQOOSTETFSMGLLCLTLFVE
ECLRRVTRKHMLIVLVITCAIILGGLTMMDLRALIMLGDWMSGRIGQIHLAIMA
VFKNSPGVGLVGRFLKRLSRETALMVLGMAMTIVLSIPHOLMELIDIGLSLGLIKTV
TOPNTQVGTIALSLTIFIRSTPLVMAWRTIMAVLFVVTILPLCRTSCLQKQSHWEI
TALLIGQAQLPVYLTMLKGNASRRSWPLNEGIMAVGLVLSLGSALLKNDVPLAGPWA
GALLAAYVMSGSADLSLEKANVQWDEMADITGSSPIIEVKODEDSFSIRDEET
NMITLLYKLA LITVSGLYPLAIPVMTLWYMWQVKTQSRGALMDVPPAPATKKALES
GVYRIMQGLGFTQVGVGHMEGVFHTMHVTRGVSIVCHETGRLEPSWADVRNDMS
YGGWRKGLDKWKEEDVQVLAIBPGKNPKHVOTKPLFKTLTGIGAVLTDLPKPGTSG
SPIINRKGVLGVLGNGVTVKSGDYVSALTOAERI GEPDVEDEDI ERKRLTMDHL
PGAGTKRILPSIVREALKRLRLTILAIPTRVAAEMBEALRGLPIRYQTPAVKSEHT
GREIVLWCHATFTFRLSSSPEDYNI VDEAHFTDPSVAAARGYIVSTRVENGEA
AIFMTATPGATDPSNPIEDIREIPERSWNTGDMITDYQKTVFVPS IKAG
NDIANCLRSKPKVILQLSRKTFTDEYPTKLTLDMDVFTVTDISEMGANFRAGRVIDR
RCLAPVTLDPGPERVILAGIPVTPASAAORRGRIRNPAQEDDQVYVFGDPLKNDSD
HAHTEAKMLDNLTPGELIIPTLFPGPERKTOAIDGEEFLRGEOR TPTVELMRGD
PVWLSKYVASNGISYKDRWCFTGERNNQILENNEVEIWTREGEKKKLAPRWLDARV
YADPMALCKFPASGRKSI TLDILTEIASLPTLYLSRAKLALDNI VMLHHTREGRRA
EIQOALNBPESLETFMLVGLIPEPERKQTPQDNQLIYVILITLITLGLIAANBMGLI
EIQPQWIAASIIIEFFLMVLIPEPERKQTPQDNQLIYVILITLITLGLIAANBMGLI
EKTCTDFGFKVETPITLDVDRPASAWTLA VATTILTPMLRHTI ENTSAULSLAA
IANQAAVLMLGKGLVHRLMDLGPVLLAMGCYSQVNP TLTASLAVMLLHVAYIIGPGL
QAKATRAQKRTAAGIMKNPTVDIGITVIDLEPISYDEPKFKQLQGVMLLVLCAGQLL
MRTTWAFCVELTLAGPILTLWEGNPGFWNTTIAVSTANI FRGSYLAGAGLAPSLIK
NAQTPRGVGTGTGLTKWKRQLNSDRKEFEYKRSGLILEVDRTSAKLDKGSKI
KHAVSRGSKIRMTIVERGMVKGVKVDLDCRGCSYMYMATLKNVTEVAKYTKGGPG
HEBPIMATYGNWLVKHSVDVFPYKPTQVDTLLCDIGESSNPTIEEGRTLVLMK
VEPWLSSKPECI KVLNPMYPTVIELEKLRQKHGGLVRCPLSRNSTHEMYVWSGAS
GNIVSSVNTTS KMLNRP TTHRKPTTEKQVDLDGATRSVSTETEKEDMTITIGRLQR
LQEBHEKTHYDQNPRTYWAYHGSYEAPSTGSASWNGVVKLLTLPWDPVI PMVTQL
AMTDTTPFGQORVFEKVDTRTPQPKPTRMVTMTANWLWALIKKKKPNLCRTREF
ISKVRSNAAICAVFOEQGWTSAEAVNDSRFWELVDKERALHQBEGKCESCVYMMCK
REKLGFRGAKSRAIYMWLGARFLEFEALGFLNEDHWFGRSNSVGEQGLHRL
GYLLEIDKQDGLMYADDITAGWDTRITEDDLQNEBELITEQMAPHHKLAKAIKFLTY
KNVVKVLRTPFRGAVMDIISRKDQSGQVGTGLNTFTNMEVOLLIRQMEABEVITQ
DMQNPGLKERVKEVKGCVGLDKRMALISGDDCVVKPLDERFGTSLFLINDMKYR
KDI POMEPSKGNWQSVFPCSHFHFI FMKDGRSLVPCRNODELLGRARISOGAGW
SLRETACLKRAYAQMWSLWYFHREDRLASMAICSAVPTWFTPTSTWTSIHAHOMW
TTEDMLKWNVRWLEDNPNMTDTPVHSWEDI PYLGRKREDLWCSLIGLSRATWAKN
IHTAITQVRNLIGKEIEVDIPVPMKRYSAFSESEGVLT"
/mat_peptide 102..440
/product="anchored capsid protein"
/mat_peptide 102..398
/product="virion capsid protein"
/mat_peptide 441..938
/nc="virc"
/mat_peptide 714..938
/product="prM"
/mat_peptide 939..2423
/nc="M"
/product="envelope protein"
/mat_peptide 1109
TGTGGAACACCTACATCTGTATTAAACGAGCATTATTGAAGCATTATTCACGGGTATTGT 1168
Db 11157 TGTGGAATCTCATACTCTCTCTTTTCAATATATTGAAGCATTATTCACGGGTATTGT 11216
Qy 11169 CTGATGCGGATACATATTTGAATGTATTTAGAAAAATAAACAATAGGGGTTCGCGC 1228
Db 11217 CTGATGCGGATACATATTTGAATGTATTTAGAAAAATAAACAATAGGGGTTCGCGC 11276
Qy 1229 ACATTTCCCGGAAAGTGCCACTGACGTCTTAAGAAACCATTTATCATGACATTAAAC 1288
Db 11277 ACATTTCCCGGAAAGTGCCACTGACGTCTTAAGAAACCATTTATCATGACATTAAAC 11336
Qy 1289 TATAAAAATAGCGGTATACAGAGCGCTTTCGTCTTCAAGAAATTCATGTTGACAGCT 1348
Db 11337 TATAAAAATAGCGGTATACAGAGCGCTTTCGTCTTCAAGAAATTCATGTTGACAGCT 11396
Qy 1349 TATCATCGATAAGCTTTAAATGCGGTAGTTTATCACAGTTAAATTTGCTAACGAGTCAGC 1408
Db 11397 TATCATCGATAAGCTTTAAATGCGGTAGTTTATCACAGTTAAATTTGCTAACGAGTCAGC 11456
Qy 1409 ACCGTGTATGAAATCTAAACAATGCGCTCATGCTCATCTCGGACCGCTCACCCCTGGATGC 1468
Db 11457 ACCGTGTATGAAATCTAAACAATGCGCTCATGCTCATCTCGGACCGCTCACCCCTGGATGC 11516
Qy 1469 TGTAGGATAGGCTTGTTATGCGGTAATGCGGCTCTTTCGGGGCTCTTTCGGGGATATCGTCCATTC 1528
Db 11517 TGTAGGATAGGCTTGTTATGCGGTAATGCGGCTCTTTCGGGGCTCTTTCGGGGATATCGTCCATTC 11576
Qy 1529 CGACAGCATCGCCAGTCACATATGCGGTGCTGCTGCGCGCTATATGCGGTGATGCAATTTCT 1588
Db 11577 CGACAGCATCGCCAGTCACATATGCGGTGCTGCTGCGCGCTATATGCGGTGATGCAATTTCT 11636
Qy 1589 ATGCGACCGCTTCTCGGACCATGTCGCGCGCTTTTGGCGCGCGCCAGCTCTGCTCGC 1648
Db 11637 ATGCGACCGCTTCTCGGACCATGTCGCGCGCTTTTGGCGCGCGCCAGCTCTGCTCGC 11696
Qy 1649 TTGCTACTTGGAGCCACTATCGACTACGGATCATGCGGACCAACACCCGCTCTGTGAT 1708
Db 11697 TTGCTACTTGGAGCCACTATCGACTACGGATCATGCGGACCAACACCCGCTCTGTGAT 11756
Qy 1709 CCTCTAGCGCGGACCATCGTGGCGGCATCACCGCGGCACACAGTGGGTGCTGGCGC 1768
Db 11757 CCTCTAGCGCGGACCATCGTGGCGGCATCACCGCGGCACACAGTGGGTGCTGGCGC 11816
Qy 1769 CTATATCGCCACATCACCGATGGGGAAGATCGGGCTCGCCACTTCGGGCTCATGAGCGC 1828
Db 11817 CTATATCGCCACATCACCGATGGGGAAGATCGGGCTCGCCACTTCGGGCTCATGAGCGC 11876
Qy 1829 TTGTTTCGGGTGGGTATGTTGGCAGGCCCGCTGGCGGGGATCTGTTGGCGGCATCTC 1888
Db 1829 TTGTTTCGGGTGGGTATGTTGGCAGGCCCGCTGGCGGGGATCTGTTGGCGGCATCTC 1888
```

| | | | |
|------------|------------|--|------------------------------|
| Db | 11877 | TTGTTTCGGGCTGATGTTGGCAGGCCCGTGGCCGGGGGACTGTTGGGGCGGCATCTC | 11936 |
| Qy | 1889 | CTTGCATGCAACATCTCTTGGCGGGGGTGGCTCAACGGGCTCAACCTACTACTTGGGCTG | 1948 |
| Db | 11937 | CTTGCATGCAACATCTCTTGGCGGGGGTGGCTCAACGGGCTCAACCTACTACTTGGGCTG | 11996 |
| Qy | 1949 | CTTCTTAATGAGAGTGCATTAAGGAGAGAGCGTCGACCGATGCCCCCTTGAGAGCCTTCAA | 2008 |
| Db | 11997 | CTTCTTAATGAGAGTGCATTAAGGAGAGAGCGTCGACCGATGCCCCCTTGAGAGCCTTCAA | 12056 |
| Qy | 2009 | CCAGTCAGCTCTCTTCCGCTGGGGGGGCGGCGGATGACTATCGTCCCGGCACTTATGACTGT | 2068 |
| Db | 12057 | CCAGTCAGCTCTCTTCCGCTGGGGGGGCGGCGGATGACTATCGTCCCGGCACTTATGACTGT | 12116 |
| Qy | 2069 | CTTCTTTATCATGCAACTCGTAGACAGGTCCGGCAGCGCTCTGGGTCACTTTTCGGCGA | 2128 |
| Db | 12117 | CTTCTTTATCATGCAACTCGTAGACAGGTCCGGCAGCGCTCTGGGTCACTTTTCGGCGA | 12176 |
| Qy | 2129 | GGACCGCTTTTCGCTGGAGCGGAGATGATCGGCTCTGCTTCGCGGTATTTTCGGAATCTT | 2188 |
| Db | 12177 | GGACCGCTTTTCGCTGGAGCGGAGATGATCGGCTCTGCTTCGCGGTATTTTCGGAATCTT | 12236 |
| Qy | 2189 | GCAGCCCTCGCTCAAGCTTCGTCACTGGTCCCGCACCAAGCTTTCGGCGGAGAGCA | 2248 |
| Db | 12237 | GCAGCCCTCGCTCAAGCTTCGTCACTGGTCCCGCACCAAGCTTTCGGCGGAGAGCA | 12296 |
| Qy | 2249 | GGCAATATCGCCGGGATGGCGGCGAGCGCTGGGCTACGTCTGTCTGGGTTTCGGGAC | 2308 |
| Db | 12297 | GGCAATATCGCCGGGATGGCGGCGAGCGCTGGGCTACGTCTGTCTGGGTTTCGGGAC | 12356 |
| Qy | 2309 | GCAGGCTGGATGCTTCCCATATGATGATCTTCTCGCTTCGGCGGCGATCGGGATGCC | 2368 |
| Db | 12357 | GCAGGCTGGATGCTTCCCATATGATGATCTTCTCGCTTCGGCGGCGATCGGGATGCC | 12416 |
| Qy | 2369 | CGCGTTGAGGCCATCTGTCTCAGGCGAGTAGATGACGACCATCAGGAGACAGCTTCAAG | 2428 |
| Db | 12417 | CGCGTTGAGGCCATCTGTCTCAGGCGAGTAGATGACGACCATCAGGAGACAGCTTCAAG | 12476 |
| Qy | 2429 | ATCGCTCGCGCTCTTACAGCCTTAACTTCATCATTTGAGCGCTGATCGTACGGCGAT | 2488 |
| Db | 12477 | ATCGCTCGCGCTCTTACAGCCTTAACTTCATCATTTGAGCGCTGATCGTACGGCGAT | 12536 |
| Qy | 2489 | TTATGCGCCCTCGGCGACATGAGACGGTTGGCATGGATTTGAGCGCGCCCTATA | 2548 |
| Db | 12537 | TTATGCGCCCTCGGCGACATGAGACGGTTGGCATGGATTTGAGCGCGCCCTATA | 12596 |
| Qy | 2549 | CTTGTCTGCTCCCGGTTGCGTTCGGTTCGATGAGCGCGGCGACCTCGACCTGAAT | 2608 |
| Db | 12597 | CTTGTCTGCTCCCGGTTGCGTTCGGTTCGATGAGCGCGGCGACCTCGACCTGAAT | 12656 |
| Qy | 2609 | GGAGCGCGGCGACCTCGCTAAACGATTCACCTCAAGAAATTGGAGCCAATCAATTC | 2668 |
| Db | 12657 | GGAGCGCGGCGACCTCGCTAAACGATTCACCTCAAGAAATTGGAGCCAATCAATTC | 12716 |
| Qy | 2669 | TTGCGGAACTGTGAATGCCCAACCAACCTTGGCAGACATATCATCGCTCGGCC | 2728 |
| Db | 12717 | TTGCGGAACTGTGAATGCCCAACCAACCTTGGCAGACATATCATCGCTCGGCC | 12776 |
| Qy | 2729 | ATCTCCAGCGCGCACCGGCGCATCTCGGCGAGCGTTGGGTCTCTGG----- | 2776 |
| Db | 12777 | ATCTCCAGCGCGCACCGGCGCATCTCGGCGAGCGTTGGGTCTCTGG----- | 12836 |
| Qy | 2777 | ----- | 2776 |
| Db | 12837 | ATGATCGTCTCTGCTGTGAGGACCGGCTAGGCTGGCGGGTTGCTTACTGTTAG | 12896 |
| Qy | 2777 | ----- | 2776 |
| Db | 12897 | CAGATGAATCACCAGATACGGAGCGAACGTGAAGCGACTGCTGTCTCAAAACGTCTCGG | 12956 |
| Qy | 2777 | ----- | 2776 |
| Db | 12957 | ACCTGAGCAACACATGAATGCTCTTGGTTTCGGTTTCGTAAGACTTGAAACGGG | 13016 |
| Qy | 2777 | ----- | 2776 |
| Db | 13017 | AAGTCAGCGCCTGCACCAATTATGTTCCGGATCTGCATCGCAGGATGCTGCTGGCTACCC | 13076 |
| Qy | 2777 | -----GCTGGCAATTGACCTCGAGTATTTTCTC | 2805 |
| Db | 13077 | TGTGAACACCTACATCTGTATTAAACGAGCGTGGCATTTGACCTTGAGTATTTTCTC | 13136 |
| Qy | 2806 | TGTCGCCCGCATCCATACCGCAGTTGTTTACCCCTCAACAGTTCCAGTAAACCGGCA | 2865 |
| Db | 13137 | TGTCGCCCGCATCCATACCGCAGTTGTTTACCCCTCAACAGTTCCAGTAAACCGGCA | 13196 |
| Qy | 2866 | TGTTTCATCATCAGTAACCCCGTATCGTGAGCATCTCTCTCTGTTTCATCGGTATCATTACC | 2925 |
| Db | 13197 | TGTTTCATCATCAGTAACCCCGTATCGTGAGCATCTCTCTCTGTTTCATCGGTATCATTACC | 13256 |
| Qy | 2926 | CCATGAACAGAAATCCCTTACACGAGGATCAGTGACCAACAGGAAAAACCGCC | 2985 |
| Db | 13257 | CCATGAACAGAAATCCCTTACACGAGGATCAGTGACCAACAGGAAAAACCGCC | 13316 |
| Qy | 2986 | CTTAACATGCGCCGCTTTATCAGAAGCCAGATTAACGTTCTGGAGAACTCAACGAG | 3045 |
| Db | 13317 | CTTAACATGCGCCGCTTTATCAGAAGCCAGATTAACGTTCTGGAGAACTCAACGAG | 13376 |
| Qy | 3046 | CTGGACCGGATGAACAGGCGAGCTCT 3073 | |
| Db | 13377 | CTGGACCGGATGAACAGGCGAGCTCT 13404 | |
| RESULT 8 | | | |
| AV376438 | | 15239 bp | DNA circular SYN 09-OCT-2003 |
| LOCUS | AV376438 | Dengue virus vector p4 (Delta30), complete sequence. | |
| DEFINITION | AV376438 | | |
| ACCESSION | AV376438 | | |
| VERSION | AV376438.1 | GI:36312945 | |
| KEYWORDS | | | |
| SOURCE | | Dengue virus vector p4 (Delta30) | |
| ORGANISM | | Dengue virus vector p4 (Delta30) | |
| REFERENCE | | other sequences; artificial sequences; vectors. | |
| AUTHORS | | 1 (bases 1 to 15239) | |
| | | Durbin, A.P., Karron, R.A., Sun, W., Vaughn, D.W., Reynolds, M.J., | |
| | | Perreault, J.R., Thumar, B., Men, R., Lai, C.J., Elkins, W.R., | |
| | | Chanock, R.M., Murphy, B.R. and Whitehead, S.S. | |
| | | Attenuation and immunogenicity in humans of a live dengue virus | |
| | | type-4 vaccine candidate with a 30 nucleotide deletion in its | |
| | | 3'-untranslated region | |
| | | Am. J. Trop. Med. Hyg. 65 (5), 405-413 (2001) | |
| JOURNAL | | 11716091 | |
| PUBMED | | | |
| REFERENCE | | 2 (bases 1 to 15239) | |
| AUTHORS | | Whitehead, S.S. | |
| TITLE | | Direct Submission | |
| JOURNAL | | Submitted Drive, Room 6515, Bethesda, MD 20892, USA | |
| | | 50 South | |
| FEATURES | | Location/Qualifiers | |
| source | | 1. 15239 | |
| | | /organism="Dengue virus vector p4 (Delta30)" | |
| | | /mol_type="other DNA" | |
| | | /db_xref="taxon:246753" | |
| | | /focus | |
| source | | 1. 10618 | |
| | | /organism="Dengue virus type 4" | |
| | | /mol_type="other DNA" | |
| | | /db_xref="taxon:11070" | |
| source | | 10628. .14989 | |
| | | /organism="Cloning vector pBR322" | |
| | | /mol_type="other DNA" | |
| | | /db_xref="taxon:47470" | |
| source | | 14950. .15239 | |
| | | /organism="Cloning vector pCEM-3" | |
| | | /mol_type="other DNA" | |
| | | /db_xref="taxon:90108" | |
| CDS | | 102. .10265 | |

[illegible]

| | | | |
|---|-------|---|-------|
| Df | 11757 | CCTACGCCGACGATCGTGGCCGCGCATCACCGCGCCACAGGTGGTGGTCTGGCGC | 11816 |
| Qy | 1769 | CTATATCGCGACATACCGATGCGGAGATCGGGCTCGCACATTCGGGGCTCATGAGCGC | 1828 |
| Df | 11817 | CTATATCGCGACATACCGATGCGGAGATCGGGCTCGCACATTCGGGGCTCATGAGCGC | 11876 |
| Qy | 1829 | TTGTTCGGCGTGGGTATGGTGGCAGGCCCGTGGCGGGGACTGTTGGCGGCCATCTC | 1888 |
| Df | 11877 | TTGTTCGGCGTGGGTATGGTGGCAGGCCCGTGGCGGGGACTGTTGGCGGCCATCTC | 11936 |
| Qy | 1889 | CTTCGATGCACTTCCTTGGCGCGGGTGGTCTCAACGGGCTCAACCTACTACTGGGGTG | 1948 |
| Df | 11937 | CTTCGATGCACTTCCTTGGCGCGGGTGGTCTCAACGGGCTCAACCTACTACTGGGGTG | 11996 |
| Qy | 1949 | CTTCCTAATGAGAGTCGATAGGAGAGAGCGTCGACCGATGCCCTTGAGAGCCTTCAA | 2008 |
| Df | 11997 | CTTCCTAATGAGAGTCGATAGGAGAGAGCGTCGACCGATGCCCTTGAGAGCCTTCAA | 12056 |
| Qy | 2009 | CCCACTGAGCTTCCTCGGTGGCGCGGGGATGACTATCGTCGCGGCACTTATGACTGT | 2068 |
| Df | 12057 | CCCACTGAGCTTCCTCGGTGGCGCGGGGATGACTATCGTCGCGGCACTTATGACTGT | 12116 |
| Qy | 2069 | CTTCCTTATCATCAACTCTGAGACAGGTGCCGCGACGCTCTGGGTCAATTTTCGGCGA | 2128 |
| Df | 12117 | CTTCCTTATCATCAACTCTGAGACAGGTGCCGCGACGCTCTGGGTCAATTTTCGGCGA | 12176 |
| Qy | 2129 | GGACCGCTTCCTGGAGCGGACGATGATCGGCTCTGGGTATTCGGAATCTT | 2188 |
| Df | 12177 | GGACCGCTTCCTGGAGCGGACGATGATCGGCTCTGGGTATTCGGAATCTT | 12236 |
| Qy | 2189 | GCAGCGCTCTCAAGCTTCCTGCTCACTGGTCCGCGCACCAAGCTTTCGGCGGAGAAGCA | 2248 |
| Df | 12237 | GCAGCGCTCTCAAGCTTCCTGCTCACTGGTCCGCGCACCAAGCTTTCGGCGGAGAAGCA | 12296 |
| Qy | 2249 | GGCCATTATCCCGGCAATGGCGCGACGCGCTGGGTACGCTTGTGTGGCGTTTCGGAC | 2308 |
| Df | 12297 | GGCCATTATCCCGGCAATGGCGCGACGCGCTGGGTACGCTTGTGTGGCGTTTCGGAC | 12356 |
| Qy | 2309 | CGGAGGTGGATGGCTTCCCATATGATTCCTCTCGCTTCGCGCGGCATCGGGATGCC | 2368 |
| Df | 12357 | CGGAGGTGGATGGCTTCCCATATGATTCCTCTCGCTTCGCGCGGCATCGGGATGCC | 12416 |
| Qy | 2369 | CGCGTTGAGCGCATGCTGTCCAGCAGGTAGATGACGACCATCAGGACAGCTTCAAGG | 2428 |
| Df | 12417 | CGCGTTGAGCGCATGCTGTCCAGCAGGTAGATGACGACCATCAGGACAGCTTCAAGG | 12476 |
| Qy | 2429 | ATCGCTCGCGCTCTTACAGCCTAACTTCGATCAATTTGACCGCTGATCGTCACGGCGAT | 2488 |
| Df | 12477 | ATCGCTCGCGCTCTTACAGCCTAACTTCGATCAATTTGACCGCTGATCGTCACGGCGAT | 12536 |
| Qy | 2489 | TTATGCGCGCTCGCGAGCACATGGAACGGGTGGCATGGATTTGAGCGCGCGCTTATA | 2548 |
| Df | 12537 | TTATGCGCGCTCGCGAGCACATGGAACGGGTGGCATGGATTTGAGCGCGCGCTTATA | 12596 |
| Qy | 2549 | CTTGTCTGCTCCCGGTTGGCTCGCGGTGATGAGCGCGGCACCTCGACTGAT | 2608 |
| Df | 12597 | CTTGTCTGCTCCCGGTTGGCTCGCGGTGATGAGCGCGGCACCTCGACTGAT | 12656 |
| Qy | 2609 | GGAGCGCGCGCACCTCGCTAAACGATTCACCACTCCAAGAAATGGAGCAATCAATTC | 2668 |
| Df | 12657 | GGAGCGCGCGCACCTCGCTAAACGATTCACCACTCCAAGAAATGGAGCAATCAATTC | 12716 |
| Qy | 2669 | TTGCGGAGAACTGTGAATCGCGAAACCAACCTTTGGCAGAACATATCCATCGGTCGCGC | 2728 |
| Df | 12717 | TTGCGGAGAACTGTGAATCGCGAAACCAACCTTTGGCAGAACATATCCATCGGTCGCGC | 12776 |
| Qy | 2729 | ATCTCCAGACCGCACCGCGGATCTCGGCGAGCGTTGGTCTGG----- | 2776 |
| Df | 12777 | ATCTCCAGACCGCACCGCGGATCTCGGCGAGCGTTGGTCTGG----- | 12836 |
| Qy | 2777 | ----- | 2776 |
| Df | 12837 | ATGATCGTGTCTCTGTTGAGGACCGGCTAGGCTGGCGGGTGGCTTACTGGTTAG | 12896 |
| Qy | 2777 | ----- | 2776 |
| Df | 12897 | CAGAATGAATACCGATACGCGAGCGAAACGTGAAGGACTGCTGCTGCAAAAACGTCTGG | 12956 |
| Qy | 2777 | ----- | 2776 |
| Df | 12957 | ACCTGAGCAACAACATGAATGGTCTCGGTTCCGTTCCGTTGTTAAAGTCTGGAAACCGG | 13016 |
| Qy | 2777 | ----- | 2776 |
| Df | 13017 | AAGTCAGCGCCCTGCACCATTAATGTTCCGGATCTGCATCGCAGGATGCTGCTGGCTACCC | 13076 |
| Qy | 2777 | ----- | 2805 |
| Df | 13077 | TGTGGAACACTATACCTGTATTAACGAAGCGCTGGCATTGACCTGAGTGAATTTTCTC | 13136 |
| Qy | 2806 | TGCTCCCGCGCATCCATACCGCAGTTGTTTACCTCAACAGTTCCAGTAACCGGGCA | 2865 |
| Df | 13137 | TGCTCCCGCGCATCCATACCGCAGTTGTTTACCTCAACAGTTCCAGTAACCGGGCA | 13196 |
| Qy | 2866 | TGTTCACTACAGTAACCGGTATCGTGAGCATCCTCTCGTTTCATCGGTATCATTACC | 2925 |
| Df | 13197 | TGTTCACTACAGTAACCGGTATCGTGAGCATCCTCTCGTTTCATCGGTATCATTACC | 13256 |
| Qy | 2926 | CCCATGAACAGAAATCCCTTTACACGAGGCGATCAGTGACAAACAGGAAAAACCGCC | 2985 |
| Df | 13257 | CCCATGAACAGAAATCCCTTTACACGAGGCGATCAGTGACAAACAGGAAAAACCGCC | 13316 |
| Qy | 2986 | CTTAACATGGCCGCTTTATCAGAAAGCAGACATTAACGTTCTGGAGAAACTCAACGAG | 3045 |
| Df | 13317 | CTTAACATGGCCGCTTTATCAGAAAGCAGACATTAACGTTCTGGAGAAACTCAACGAG | 13376 |
| Qy | 3046 | CTGAGCGCGGATGAACAGGCGAGCTCT 3073 | |
| Df | 13377 | CTGAGCGCGGATGAACAGGCGAGCATCT 13404 | |
| RESULT 9 | | | |
| AY656168 | | | |
| LOCUS | | | |
| DEFINITION | | | |
| ACCESSION | | | |
| VERSION | | | |
| KEYWORDS | | | |
| SOURCE | | | |
| ORGANISM | | | |
| REFERENCE | | | |
| AUTHORS | | | |
| TITLE | | | |
| JOURNAL | | | |
| PUBMED | | | |
| AUTHORS | | | |
| TITLE | | | |
| JOURNAL | | | |
| FEATURES | | | |
| source | | | |
| 1. 15256 | | | |
| /organism="Chimeric dengue virus vector | | | |
| p4(delta30)-D3L-ME" | | | |
| /mol_type="other DNA" | | | |
| /db_xref="taxon:284998" | | | |
| /focus | | | |
| 1. 401 | | | |
| /organism="Dengue virus type 4" | | | |
| /mol_type="other DNA" | | | |
| /db_xref="taxon:11070" | | | |
| 402. .2339 | | | |
| /organism="Dengue virus type 3" | | | |
| source | | | |
| 15256 bp DNA circular SYN 14-JAN-2005 | | | |
| Chimeric dengue virus vector p4(delta30)-D3L-ME, complete sequence. | | | |
| AY656168 | | | |
| Chimeric dengue virus vector p4(delta30)-D3L-ME | | | |
| other sequences; artificial sequences; vectors. | | | |
| 1 (bases 1 to 15256) | | | |
| Blaney, J.E. Jr., Hanson, C.T., Firestone, C.V., Hanley, K.A., | | | |
| Murphy, B.R., and Whitehead, S.S. | | | |
| Genetically modified, live attenuated dengue virus type 3 vaccine | | | |
| candidates | | | |
| Am. J. Trop. Med. Hyg. 71 (6), 811-821 (2004) | | | |
| 15642976 | | | |
| 2 (bases 1 to 15256) | | | |
| Whitehead, S.S. | | | |
| Direct Submission | | | |
| Submitted (17-JUN-2004) Laboratory of Infectious Diseases, NIAID, | | | |
| Drive, Bldg 50, Room 6515, Bethesda, MD 20892, USA | | | |
| 50 South Location/Qualifiers | | | |
| 1. 15256 | | | |
| /organism="Chimeric dengue virus vector | | | |
| p4(delta30)-D3L-ME" | | | |
| /mol_type="other DNA" | | | |
| /db_xref="taxon:284998" | | | |
| /focus | | | |
| 1. 401 | | | |
| /organism="Dengue virus type 4" | | | |
| /mol_type="other DNA" | | | |
| /db_xref="taxon:11070" | | | |
| 402. .2339 | | | |
| /organism="Dengue virus type 3" | | | |

| | | | |
|----------------------------|---|--|----------------------|
| | /mol_type="other DNA" | | |
| | /db_xref="taxon:11069" | | |
| source | 2363..10635 | | |
| | /organism="Dengue virus type 4" | | |
| | /mol_type="other DNA" | | |
| | /db_xref="taxon:11070" | | |
| source | 10645..15006 | | |
| | /organism="Cloning vector pBR322" | | |
| | /mol_type="other DNA" | | |
| | /db_xref="taxon:47470" | | |
| source | 15007..15256 | | |
| | /organism="Cloning vector pGEM-3" | | |
| | /mol_type="other DNA" | | |
| | /db_xref="taxon:90108" | | |
| misc_feature | 102..10279 | | |
| | /note="encodes polyprotein precursor once XhoI linker sequence is removed prior to preparation of infectious transcripts corresponding to the viral genome" | | |
| misc_feature | 2340..2362 | | |
| | /note="XhoI linker; required for stable replication in Escherichia coli" | | |
| misc_feature | 10636..10644 | | |
| | /note="KpnI/AgeI linker" | | |
| promoter | 15238..15256 | | |
| | /note="SP6" | | |
| ORIGIN | | | |
| Query Match | 34.0%; | Score 1636.8; | DB 11; Length 15256; |
| Best Local Similarity | 86.4%; | Pred. No. 0; | |
| Matches 1943; Conservative | 0; | Mismatches 22; | Indels 283; Gaps 1; |
| Qy | 1109 | TGTGAACACCTACATCTGTATTACGAAGCAATTATTGAAGCAATTTATCAGGGTTATTGT | 1168 |
| Db | 11174 | TGTGAATACTCATACTCTCTCTTTTCAATATTATTGAAGCAATTTATCAGGGTTATTGT | 11233 |
| Qy | 1169 | CTCATGAGCGGATACATATTTGAATGTATTAGAAAAATAACAAATAGGGTTCCGGC | 1228 |
| Db | 11234 | CTCATGAGCGGATACATATTTGAATGTATTAGAAAAATAACAAATAGGGTTCCGGC | 11293 |
| Qy | 1229 | ACATTTCCCGAAAGTGCCACCTGACGCTTAAGAAACCAATTTATCATGACATTAAAC | 1288 |
| Db | 11294 | ACATTTCCCGAAAGTGCCACCTGACGCTTAAGAAACCAATTTATCATGACATTAAAC | 11353 |
| Qy | 1289 | TATAAATAAGCGGTATCACGAGGCCCTTTCTGCTTTCAAGAAATTTCTATGTTGAAGCT | 1348 |
| Db | 11354 | TATAAATAAGCGGTATCACGAGGCCCTTTCTGCTTTCAAGAAATTTCTATGTTGAAGCT | 11413 |
| Qy | 1349 | TATCATCGATAAGCTTTAATCGGTTAGTTTATCACAGTTTAAATTTGCTAACCGAGTCAGGC | 1408 |
| Db | 11414 | TATCATCGATAAGCTTTAATCGGTTAGTTTATCACAGTTTAAATTTGCTAACCGAGTCAGGC | 11473 |
| Qy | 1409 | ACCGTGTATGAATCTTAACAATGCGCTCATCGTCTATCTCGGCACCGTCACCTGGATGC | 1468 |
| Db | 11474 | ACCGTGTATGAATCTTAACAATGCGCTCATCGTCTATCTCGGCACCGTCACCTGGATGC | 11533 |
| Qy | 1469 | TGTAGGCATAGGCTTGGTTATGCGGTTACTGCCGGGCTCTTTGCGGGAATCGTCCATTTC | 1528 |
| Db | 11534 | TGTAGGCATAGGCTTGGTTATGCGGTTACTGCCGGGCTCTTTGCGGGAATCGTCCATTTC | 11593 |
| Qy | 1529 | CGACAGCATCGCCAGTCACATATGGCGTCTGCTAGCGCTATATGGGTTGATGCAATTTCT | 1588 |
| Db | 11594 | CGACAGCATCGCCAGTCACATATGGCGTCTGCTAGCGCTATATGGGTTGATGCAATTTCT | 11653 |
| Qy | 1589 | ATGCGCACCCGTTCTCGGAGCACTGTCGACCGCTTTGGCCGCGCCGACAGTCTGCTCGC | 1648 |
| Db | 11654 | ATGCGCACCCGTTCTCGGAGCACTGTCGACCGCTTTGGCCGCGCCGACAGTCTGCTCGC | 11713 |
| Qy | 1649 | TTGCTACTTTGAGCCACTTATCGACTACCGGATCATGGCGACCAACACCGCTCTGTGGAT | 1708 |
| Db | 11714 | TTGCTACTTTGAGCCACTTATCGACTACCGGATCATGGCGACCAACACCGCTCTGTGGAT | 11773 |
| Qy | 1709 | CCTCTACGCGGACGCAATCGTGGCGGCAATCACCGCGCGCAACAGGTGCGGTTGCTGGCGC | 1768 |

| | | |
|----|--|-------|
| Db | 11774 CCTCTACCGCGACGATCGTGGCCGGCATCACCGCGCGCACAGGTGCGGTTGCTGGCGC | 11833 |
| Qy | 1769 CTATATCGCCGACATCACCGATGGGGAAGATCGGCTCGCCACTTTCGGGCTCATGAGCGC | 1828 |
| Db | 11834 CTATATCGCCGACATCACCGATGGGGAAGATCGGCTCGCCACTTTCGGGCTCATGAGCGC | 11893 |
| Qy | 1829 TTGTTTCGGCGTGGGTATGTTGGCAGGCCCGTGGCCGGGGGACATGTTGGGGCGCATCTC | 1888 |
| Db | 11894 TTGTTTCGGCGTGGGTATGTTGGCAGGCCCGTGGCCGGGGGACATGTTGGGGCGCATCTC | 11953 |
| Qy | 1889 CTTGCATCACCATTTCTTTGGCGGCGGTGCTCAACGGCCTCAACCTACTACTACTGCGGCTG | 1948 |
| Db | 11954 CTTGCATCACCATTTCTTTGGCGGCGGTGCTCAACGGCCTCAACCTACTACTACTGCGGCTG | 12013 |
| Qy | 1949 CTTCTAATGAGGAGTGCATTAAGGAGAGCGTGCAGCGATGCGCTTTGAGAGCGCTTCAA | 2008 |
| Db | 12014 CTTCTAATGAGGAGTGCATTAAGGAGAGCGTGCAGCGATGCGCTTTGAGAGCGCTTCAA | 12073 |
| Qy | 2009 CCCAGTCACTCTCTTCGGTGGCGCGGGGATGATCTATGTCGCGGCACTTATGACTGT | 2068 |
| Db | 12074 CCCAGTCACTCTCTTCGGTGGCGCGGGGATGATCTATGTCGCGGCACTTATGACTGT | 12133 |
| Qy | 2069 CTTCTTATCATGCAACTCGTAGGACAGTGCAGCGGCTCTGGGTCACTTTTCGGCGA | 2128 |
| Db | 12134 CTTCTTATCATGCAACTCGTAGGACAGTGCAGCGGCTCTGGGTCACTTTTCGGCGA | 12193 |
| Qy | 2129 GGACCGCTTTCGCTGGAGCGGACGATGATCGGCTCTGCTTGGGTGATTCGGGAATCTT | 2188 |
| Db | 12194 GGACCGCTTTCGCTGGAGCGGACGATGATCGGCTCTGCTTGGGTGATTCGGGAATCTT | 12253 |
| Qy | 2189 GCAGCCCTCGCTCAAGCCTTTCGTCACCTGGTCCCGCCACCAACGCTTTTCGGCGAGA | 2248 |
| Db | 12254 GCAGCCCTCGCTCAAGCCTTTCGTCACCTGGTCCCGCCACCAACGCTTTTCGGCGAGA | 12313 |
| Qy | 2249 GGCCATTATCGCGGCGATGGCGCGGACGCGCTGGGCTACGCTTTCGCTGGCGGTTCGCGAC | 2308 |
| Db | 12314 GGCCATTATCGCGGCGATGGCGCGGACGCGCTGGGCTACGCTTTCGCTGGCGGTTCGCGAC | 12373 |
| Qy | 2309 CGGAGGCTGGATGGCTTTCGCCATTTATGATTTCTTCGCTTTCGCGCGGATCGGATGCC | 2368 |
| Db | 12374 CGGAGGCTGGATGGCTTTCGCCATTTATGATTTCTTCGCTTTCGCGCGGATCGGATGCC | 12433 |
| Qy | 2369 CGGCTGCGAGCCATGCTGCCAGCAGGTAGATGACGACCATCAGGACAGCTTCAAGG | 2428 |
| Db | 12434 CGGCTGCGAGCCATGCTGCCAGCAGGTAGATGACGACCATCAGGACAGCTTCAAGG | 12493 |
| Qy | 2429 ATCGCTCGCGCTCTTACCAGCCTAACTTCGATCATTTGGAACGCTGATCGTCAAGCGAT | 2488 |
| Db | 12494 ATCGCTCGCGCTCTTACCAGCCTAACTTCGATCATTTGGAACGCTGATCGTCAAGCGAT | 12553 |
| Qy | 2489 TTATCGCGCTCGCGGACATGGAACGCGTTCGSCATGGATTGTAGGCGCGCCCTATA | 2548 |
| Db | 12554 TTATCGCGCTCGCGGACATGGAACGCGTTCGSCATGGATTGTAGGCGCGCCCTATA | 12613 |
| Qy | 2549 CTTTGTCTGCTCCCGGTTGCGTTCGCGGTGSCATGGAGCGCGGCGCACTCGACCTGAAT | 2608 |
| Db | 12614 CTTTGTCTGCTCCCGGTTGCGTTCGCGGTGSCATGGAGCGCGGCGCACTCGACCTGAAT | 12673 |
| Qy | 2609 GGAAGCGCGGCGACCTCGCTAAACGGAATTAACCTCAAGAAATTTGGAGCAATCAATTC | 2668 |
| Db | 12674 GGAAGCGCGGCGACCTCGCTAAACGGAATTAACCTCAAGAAATTTGGAGCAATCAATTC | 12733 |
| Qy | 2669 TTGCGGAGAACTGTGAATGGCAAAACCAACCTTTGGCAGAACATATCATCGCTCGGCC | 2728 |
| Db | 12734 TTGCGGAGAACTGTGAATGGCAAAACCAACCTTTGGCAGAACATATCATCGCTCGGCC | 12793 |
| Qy | 2729 ATCTCCAGCAGCCGCGCGGATCTCGGGCAGCGTTCGGGTCTCTGG----- | 2776 |
| Db | 12794 ATCTCCAGCAGCCGCGCGGATCTCGGGCAGCGTTCGGGTCTCTGGCCACCGGTGCGC | 12853 |
| Qy | 2777 ----- | 2776 |

[illegible]

| | | | |
|------------|---|---|------------------------------|
| Qy | 2777 | ----- | 2776 |
| Ddb | 12866 | ATGATCGTGCTCCTGTCTGTTGAGGACCGGCTAGGCTGGCGGTTGCTTACTGGTTAG | 12925 |
| Qy | 2777 | ----- | 2776 |
| Ddb | 12926 | CAGATGAATCACCGATACGCGAGCGAAGCTGAAGCGACTGCTGCTCAAAACGTCGCG | 12985 |
| Qy | 2777 | ----- | 2776 |
| Ddb | 12986 | ACCTGAGCAACAACATGAATGCTCTTCGGTTTCGGTGTTCGTAAGTCTGGAACCGGG | 13045 |
| Qy | 2777 | ----- | 2776 |
| Ddb | 13046 | AGTTCAGCGCCCTGCACCATTAATGTTCCGGATCTGCATCGCAGGATGCTGCTGGCTACCC | 13105 |
| Qy | 2777 | -----GCTGGCATTGACCTGAGTGAATTTTCTC | 2805 |
| Ddb | 13106 | TGTGGAACACCTACATCTGTATTAAAGCGCTGGCATTCACCTGAGTGAATTTTCTC | 13165 |
| Qy | 2806 | TGTCCTCCGCGCATCCATACCGCAGTTGTTTACCCCTCACACGTTCCAGTAACCGGGCA | 2865 |
| Ddb | 13166 | TGGTCCGCGCATCCATACCGCCAGTTGTTTACCCCTCACACGTTCCAGTAACCGGGCA | 13225 |
| Qy | 2866 | TGTTTCATCATCAGTAACCCGATCTGTGAGCATCTCTCTCGTTTCATCGGTATCATTTACC | 2925 |
| Ddb | 13226 | TGTTTCATCATCAGTAACCCGATCTGTGAGCATCTCTCTCGTTTCATCGGTATCATTTACC | 13285 |
| Qy | 2926 | CCCATGAACAGAAATCCCCCTTACACGAGGAGCATCAGTGACCAACAGAAAAACCGCC | 2985 |
| Ddb | 13286 | CCCATGAACAGAAATCCCCCTTACACGAGGAGCATCAGTGACCAACAGAAAAACCGCC | 13345 |
| Qy | 2986 | CTTAACATGGCCCGCTTTATCAGAGCCAGACATTAACGCTTCTCGAGAACTCAACGAG | 3045 |
| Ddb | 13346 | CTTAACATGGCCCGCTTTATCAGAGCCAGACATTAACGCTTCTCGAGAACTCAACGAG | 13405 |
| Qy | 3046 | CTGAGCGCGGATGAACAGCGAGACTCT | 3073 |
| Ddb | 13406 | CTGAGCGCGGATGAACAGCGAGACTCT | 13433 |
| RESULT 11 | | | |
| LOCUS | AY243466 | 15270 bp | DNA circular SYN 20-OCT-2003 |
| DEFINITION | Chimeric Dengue virus vector p4-D2-ME, complete sequence. | | |
| ACCESSION | AY243466 | | |
| VERSION | AY243466.1 | GI:30026599 | |
| KEYWORDS | | | |
| SOURCE | Chimeric Dengue virus vector p4-D2-ME | | |
| ORGANISM | Chimeric Dengue virus vector p4-D2-ME | | |
| REFERENCE | Other sequences; artificial sequences; vectors. | | |
| AUTHORS | 1 (bases 1 to 15270) Elkins,W.R. and Murphy,B.R. | | |
| TITLE | Substitution of the structural genes of dengue virus type 4 with those of type 2 results in chimeric vaccine candidates which are attenuated for mosquitoes, mice, and rhesus monkeys | | |
| JOURNAL | Vaccine 21 (27-28), 4307-4316 (2003) | | |
| PUBMED | 14505913 | | |
| REFERENCE | 2 (bases 1 to 15270) | | |
| AUTHORS | Whitehead,S.S. | | |
| TITLE | Direct Submission | | |
| JOURNAL | Submitted (24-FEB-2003) LID, NIAID, 50 South Dr., Room 6515, Bethesda, MD 20892, USA | | |
| FEATURES | Location/Qualifiers | | |
| source | 1..15270 | | |
| | /organism="Chimeric Dengue virus vector p4-D2-ME" | | |
| | /mol_type="genomic DNA" | | |
| | /db_xref="taxon:227155" | | |
| | /focus | | |
| | 1..404 | | |
| source | /organism="Dengue virus type 4" | | |

/mol_type="genomic DNA"
/db_xref="taxon:11070"
405..2345
/organism="Dengue virus type 2 (NGC-prototype)"
/mol_type="genomic DNA"
/db_xref="taxon:11065"
2346..10649
/organism="Dengue virus type 4"
/mol_type="genomic DNA"
/db_xref="taxon:11070"
102..10265
/codon_start=1
/product="polyprotein precursor"
/protein_id="AAP06252.1"
/db_xref="GI:30026600"
/translation="MNQRKKVRRPPNMLKRNRVSTPQGLVKRFRSTGLPSGKGLR
MVLATITFLVLSITTAGILKRWGLKKNKAIKILGFRKEIGRMNLNLRKRSAG
MILMLITVMAFHLITTRNGPHMIVSRQEKGLSLLFTEDGVNCTLMAMDGLCE
TITVCPKLRQNEPEDIQCNSTSTVYTGCTTGEHREKRSVALPHVHGML
RTTWSSEGAKHAORIETWILRHFGFTTMAALAYTIGTTHFORALIFLLTAVAP
SMTMRCIGIENRDPVEGSGSVDIIVLEHSGCVTTMAKNPITLDFELIKTEAKQAT
LRKCIERAKLITNTTESRCTQGEPSINEBQDKRFVCKHSMVDRGNGCGLFGKGGI
VTCAFTCKKMEGVVQENLETTIVITPHSGEHAAGNDTGHGKEIKITPQSSIT
EAEITGTYTFTKPHAKQDVVVLGQEGAMHTALTATGTEIQMSGNLLFTGHKLCRL
NMIOKETLVITFKPHAKQDVVVLGQEGAMHTALTATGTEIQMSGNLLFTGHKLCRL
RMDKLOLKGMSYCMCTGKFKVVKIEAETQHGTVIIVQVEGDSGPKIPEIMLEKR
HVLGRLLTVNPIVTEKDSPVNIEAEPFGDSYIIIGVEPGOLKNWFKKSSIGOMFE
TMRGAKMAILGDTANDFSLGGVFTSIGKALHOVFGAIYGAAPSGVSWTMKILIGV
IITWGNRNTSMWATCIAVGITLFLGFTVQADMCGCVASGKELKCGSGIFVVDN
VHTWTEYKQPSPARLALNAHDKGCGVIRHSTRLNENMMKQIINENYVLMWG
SHDITVAVAGVGLTKRALTTPVSDLYKSWTKGKAKIPTPEARNSTFLIDPDT
SECNERNRWNLSLEVDEYFGMTNTIWMKPREGSEVCDHRLMSAIIKDGKAVHAD
MYLESKNTQWOLEKASLIVTKLWPKTHLWSNGVLESOMLI PKYAGPFSOHNY
RQYATQVGPWHGLKLEIDFGCEPCTTIOEDCDHRGPSLRTTASGLKLTQWCCR
SCMPPRLFGEDGCTGMEIRPLSEKEENMKVSKQVTAQGTSETFSNGLLCLTAVE
ECLRRVTRVIRHMLVVTICAILLGLTWMDLRALIMUGDWTGSRGTLQGLHLAIMA
VFKRSPGVVLGFLRLTSRETLMTVMAMTTLVSI PHDLMELIDIGISGLILKLV
TOPNTQVGLALSLTIFRSTMPVMAWRTIMAVFVVTIPLCRTSCLOQSHWEI
TALILGAQALPVYLTLMKGASRSPWLNIEGIMAVGLVSLGALLKNDVPLAGMVA
GOLLAAYVMSGSSADLSLEKAAVQWDEMADITGSSPI IEVKODEDPSPIRDEET
NMITLLVKALITVSGSLPLAI PVMTILWYMWQVKTQBSGALNDVPSAKKALSE
GVYIMQRLFGKTVGVGHMEGVFHTMHVTRGVSICHTETRLSPSWADVNDMIS
YGGWRLGDKWDEEDVQLAIEPGKPKHVQTKRPLKTLTGIGAVTLDFKPTSG
SPIINRKGKVLGNGVTVKSGDYSAITQAEIRIGEPDEVEDIEFRKRLTIMDLH
PGAKTALPSIIVREALKRLRLTILAPTRVAAAEALRGPLIRYQTPAVSEGA
GREIVDLCHATFTRLLSSTPVPYMLIIVMDEAHTDPSSVAARGYISTVMECEAA
AIFMTATPGATDPPQSNSEIEDIERIPERSWNTGFDWITDYQKTVWVPSIKAG
NDIANLRKSGKKVIQSRKFTDTEYPKTLTDMDFVVTITDI SEMGANFRAGRVIDPR
RCLKPVILPDGPERVILAGPIPVTPASAAQRRIGRNPAGDQDDQYVFGDPLKND
HAHWTEAKMLDNIYTPBGI IPTLFGPERKTOAIDGFRLEGRKKT FVELMRDGL
PVWLSYKASAGISYKOREWCTGERNNQIILEENVEIWTREGKKLRPRWLDARV
YADPMALKDFEFASGRKSTILDLITELASLPTVLSRPAKALDNI VMLHTERGERA
YOHALNELPSELTLMVLVALLGAMTAGITFLFMQKGKGLKSLMGLITIVASGLLWA
BIQPMVAASIIIEFLMLLIPEPEKORTQDNOLIYVILITLITIGLIANENGLI
TANQAAVLGMLGKGPLHRMDLGPLAMGCVSQVNTPTLTASMLLVHVAIIGPGL
IKAQTFDGYQVKTETITLDVLPASANTLYAVATITLTMLRHTIENTSANSLLA
IQAQATREAKTAAKIMKNTVDGITVDLBPISYDPRKQOLGQMLLVLCAGOLL
MRTTAFCEVILATGPIILTWEGNPGSPFWMTIIVASTANI PRGSYLAGAGLAFSLIK
NAQTPRRGTGTTGLGKWRKQNLNDRKFEFEYKRSGLIEVDTEAKSLKDGSKI
KHAVSRGSSKIRWIVERGMVPRKGVLDLGGCGGWSYIMATLKNVTEYKGTGKGGP
HEEPIPMATYGNWILVKHSGVDVFKYQEQVDTLLCDIGESSNNTIIEBGRVLRVKM
VEPILSSKPEFCIKVLPYMPYTVIELEKLRQKHGGLNVRCPPLSNSTHEMYVWSGAS
GNIVSSVNTTCKMLNLRFTTHRKPTYEKDVLGAGTSVSTETEKPDWITICRRQLR
LOEBHEKTHWDOENPYTMAVHGSYEAPSTGSSASSMVNGVVKLTAKPDWITPMTREF
AMWDTTPFGQVFEKRVDTTRTPQKPGTRWMTTANWALLQKCKESCSCYNNMGK
ISKVSNAAGAVFOEQQWTSASEAVDSRPFELVDRKALHQLGCKESCSCYNNMGK
REKKLGEFGKGRSARIWYMLGARFLEFALGFLEHDFWFGRENSWSVEGEGHLRL
GYLBEIDKQDGLMYADDTAGWDTRIITEDDILQNEELITEQWAPHKHILAKAIFKLY
QNKVVKVLRPTPRGAVMDI ISRKQDQSGQVGYTGLNTPTNNEVOLIRQMEAEVITQ
DDMQNPGLKERVKWLKECQVDRLKMAISGDDCVKPLDFERFSTLLFLNDMGKVR
KDIPOWEPFSKGNWQVFPFCSHHPHKIFMKDGRSLVVPQRNDELIGPARISQAGWM
SLRETAGLGAQAQWMSLWYFHRDLRLASMAICSAPTEWFPTRTSTWIIHAGQWM
TTEDMLKVMNRVYIENEDNPNMTDKTPVHSWEDIPYLGRKREDLWCGSLIGLSRATWAKN

source

source

CDS

IHTALTQVRNLIGKEEYVDYMPVKRYSAPSESGVL"

/product="anchored capsid protein"

/note="anchC"

102..398

/product="virion capsid protein"

/note="virC"

441..938

/product="membrane precursor protein"

/note="prM"

714..938

/product="membrane protein"

/note="M"

939..2423

/product="envelope protein"

/note="E"

2424..3479

/product="NS1 protein"

3480..4133

/product="NS2A protein"

4134..4523

/product="NS2B protein"

4524..6377

/product="NS3 protein"

6378..6758

/product="NS4A protein"

6759..6827

/product="NS4B protein"

6828..7562

/product="NS5 protein"

7563..10262

/product="NS5 protein"

10263..10262

/product="NS5 protein"

10263..10262

/product="NS5 protein"

10263..10262

/product="NS5 protein"

10263..10262

/product="NS5 protein"

10263..10262

/product="NS5 protein"

10263..10262

/product="NS5 protein"

10263..10262

/product="NS5 protein"

10263..10262

/product="NS5 protein"

10263..10262

/product="NS5 protein"

10263..10262

/product="NS5 protein"

10263..10262

/product="NS5 protein"

10263..10262

/product="NS5 protein"

10263..10262

/product="NS5 protein"

10263..10262

/product="NS5 protein"

10263..10262

/product="NS5 protein"

10263..10262

/product="NS5 protein"

10263..10262

/product="NS5 protein"

10263..10262

/product="NS5 protein"

10263..10262

/product="NS5 protein"

10263..10262

/product="NS5 protein"

10263..10262

/product="NS5 protein"

10263..10262

/product="NS5 protein"

10263..10262

ORIGIN

Query Match 34.0%; Score 1636.8; DB 11; Length 15270;
Best Local Similarity 86.4%; Pred. No. 0;
Matches 1943; Conservative 0; Mismatches 22; Indels 283; Gaps 1;

| | | | |
|----|-------|--|-------|
| Qy | 1109 | TGTGAACACCTACATCTGTATTACGAAGACATTTACGAAGATTTATCAGGGTTATTGT | 1168 |
| Db | 11188 | TGTTGAATACATCACTCTCTCTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGT | 11247 |
| Qy | 1169 | CTCATGACGGATACATATTTGAATGTATTATAGAAAAATAAACAATAGGGGTTCCGGC | 1228 |
| Db | 11248 | CTCATGACGGATACATATTTGAATGTATTATAGAAAAATAAACAATAGGGGTTCCGGC | 11307 |
| Qy | 1229 | ACATTTCCCGAAAGTCCACCTGACCTCTAAGAAACCATTTATCATGACATTAAC | 1288 |
| Db | 11308 | ACATTTCCCGAAAGTCCACCTGACCTCTAAGAAACCATTTATCATGACATTAAC | 11367 |
| Qy | 1289 | TATAAAATAGCGGTATCACGAGGCCCTTTCGTCTTCAAGAAATCTCATGTTTGACAGCT | 1348 |
| Db | 11368 | TATAAAATAGCGGTATCACGAGGCCCTTTCGTCTTCAAGAAATCTCATGTTTGACAGCT | 11427 |
| Qy | 1349 | TATCATCATGAAGCTTTAATCGGGTAGTTTATCATGATTAAATTCGTAACGAGTCAGGC | 1408 |
| Db | 11428 | TATCATCATGAAGCTTTAATCGGGTAGTTTATCATGATTAAATTCGTAACGAGTCAGGC | 11487 |
| Qy | 1409 | ACCGTGTATGAATCTACAAATGCGCTCATGCTCATCTCCGCAACCGTCACCTGGATGC | 1468 |
| Db | 11488 | ACCGTGTATGAATCTACAAATGCGCTCATGCTCATCTCCGCAACCGTCACCTGGATGC | 11547 |
| Qy | 1469 | TGTAGGCATAGGCTTGGTTATCCGGTACTCCGGGCCCTTTCGGGGATATCGTCCATTTC | 1528 |
| Db | 11548 | TGTAGGCATAGGCTTGGTTATCCGGTACTCCGGGCCCTTTCGGGGATATCGTCCATTTC | 11607 |
| Qy | 1529 | CGACAGCATCGCAGTCACTATGCGGTGCTGCTAGCGCTATATGCGTTGATGCAATTTCT | 1588 |
| Db | 11608 | CGACAGCATCGCAGTCACTATGCGGTGCTGCTAGCGCTATATGCGTTGATGCAATTTCT | 11667 |
| Qy | 1589 | ATGCGCACCCGCTTTCGAGGACACTGTCCGACCGCTTTGGCCGCCCGCCAGTCTGCTCGC | 1648 |
| Db | 11668 | ATGCGCACCCGCTTTCGAGGACACTGTCCGACCGCTTTGGCCGCCCGCCAGTCTGCTCGC | 11727 |

| | | | |
|----|-------|---|-------|
| Qy | 1649 | TTGCTACTTGGAGCCACTATCGACTACCGCATCATGCGACCAACACCGCTCTGTGGAT | 1708 |
| Db | 11728 | TTGCTACTTGGAGCCACTATCGACTACCGCATCATGCGACCAACACCGCTCTGTGGAT | 11787 |
| Qy | 1709 | CCTCTACCGCCGACGATCGTGGCCGGCATCACCGCGCCACAGGTGGGTTCGTGGCG | 1768 |
| Db | 11788 | CCTCTACCGCCGACGATCGTGGCCGGCATCACCGCGCCACAGGTGGGTTCGTGGCG | 11847 |
| Qy | 1769 | CTATATCGCCGACATCACCGATGGGGAAGATCGGCTCGCCACTTCGGGCTCATGAGCG | 1828 |
| Db | 11848 | CTATATCGCCGACATCACCGATGGGGAAGATCGGCTCGCCACTTCGGGCTCATGAGCG | 11907 |
| Qy | 1829 | TTGTTTCGGGTGGGTATGGTGGCAGGCCCGTGGCCGGGAGACTGTGTGGGGCCCATCTC | 1888 |
| Db | 11908 | TTGTTTCGGGTGGGTATGGTGGCAGGCCCGTGGCCGGGAGACTGTGTGGGGCCCATCTC | 11967 |
| Qy | 1889 | CTTCGATGACCATTCCTTGGCGGGCGGTCTCAACCGCTCAACCTACTACTCTGGCTG | 1948 |
| Db | 11968 | CTTCGATGACCATTCCTTGGCGGGCGGTCTCAACCGCTCAACCTACTACTCTGGCTG | 12027 |
| Qy | 1949 | CTTCCTAATGAGGAGTCCGATTAAGGGAGAGCGTCCGATGCGCTTTCGAGAGCTTCAA | 2008 |
| Db | 12028 | CTTCCTAATGAGGAGTCCGATTAAGGGAGAGCGTCCGATGCGCTTTCGAGAGCTTCAA | 12087 |
| Qy | 2009 | CCAGTCACTCTTTCGGTGGCGGGCGCATGATATCGTTCGCGCCACTTTATGACTGT | 2068 |
| Db | 12088 | CCAGTCACTCTTTCGGTGGCGGGCGCATGATATCGTTCGCGCCACTTTATGACTGT | 12147 |
| Qy | 2069 | CTTCTTTATCATGCACTCTGTTAGGACAGGTCCGCGGAGGCTCTGGGTCAATTTTCGCG | 2128 |
| Db | 12148 | CTTCTTTATCATGCACTCTGTTAGGACAGGTCCGCGGAGGCTCTGGGTCAATTTTCGCG | 12207 |
| Qy | 2129 | GGACCGCTTTCGCTGGAGCGGACGATGATCGGCTGTCGCTTTCGGGTATTCGGAATCTT | 2188 |
| Db | 12208 | GGACCGCTTTCGCTGGAGCGGACGATGATCGGCTGTCGCTTTCGGGTATTCGGAATCTT | 12267 |
| Qy | 2189 | GCACGCCCTCGCTCAAGCCTTTCGTCACCTGTCCTCGCCACCAACGTTTCGGCGAGA | 2248 |
| Db | 12268 | GCACGCCCTCGCTCAAGCCTTTCGTCACCTGTCCTCGCCACCAACGTTTCGGCGAGA | 12327 |
| Qy | 2249 | GGCATTTATCGCGCGCATGGCGCGGACGCGCTGGGCTACGCTTTCGCTGGGCTTCGCG | 2308 |
| Db | 12328 | GGCATTTATCGCGCGCATGGCGCGGACGCGCTGGGCTACGCTTTCGCTGGGCTTCGCG | 12387 |
| Qy | 2309 | CGAGGCTGAGTGCCTTTCCTCCCATTTATGATTTCTTCTCGCTTCGGCGGATCGG | 2368 |
| Db | 12388 | CGAGGCTGAGTGCCTTTCCTCCCATTTATGATTTCTTCTCGCTTCGGCGGATCGG | 12447 |
| Qy | 2369 | CGGTTGCGAGCCATGCTGTCTCAGGCGAGTATGATGACGACCATCAGGAGCAGCTTCA | 2428 |
| Db | 12448 | CGGTTGCGAGCCATGCTGTCTCAGGCGAGTATGATGACGACCATCAGGAGCAGCTTCA | 12507 |
| Qy | 2429 | ATCGCTCGGGCTCTTACAGCCCTAACTTCGATCAATTTGGACCGCTGATCGTCGCG | 2488 |
| Db | 12508 | ATCGCTCGGGCTCTTACAGCCCTAACTTCGATCAATTTGGACCGCTGATCGTCGCG | 12567 |
| Qy | 2489 | TTATGCGGCTTCGGCGAGCAGTGAAGCGGCTGCGATGATTTGAGCGCGCGCCCTATA | 2548 |
| Db | 12568 | TTATGCGGCTTCGGCGAGCAGTGAAGCGGCTGCGATGATTTGAGCGCGCGCCCTATA | 12627 |
| Qy | 2549 | CCTTGTCTGCTCCCGGCTTTCGCTGCGGTGCGATGAGCGCGGCGCCTCGACCTGAAT | 2608 |
| Db | 12628 | CCTTGTCTGCTCCCGGCTTTCGCTGCGGTGCGATGAGCGCGGCGCCTCGACCTGAAT | 12687 |
| Qy | 2609 | GGAGCGCGGCGCCTCTGCTAAAGGATTAACCATTCGATCAAGAAATTTGAGGCAATCA | 2668 |
| Db | 12688 | GGAGCGCGGCGCCTCTGCTAAAGGATTAACCATTCGATCAAGAAATTTGAGGCAATCA | 12747 |
| Qy | 2669 | TTGCGGAGAACTGTGAATGCGCAACCAACCTTTGGGAGACATATCCATCGCTCGCC | 2728 |
| Db | 12748 | TTGCGGAGAACTGTGAATGCGCAACCAACCTTTGGGAGACATATCCATCGCTCGCC | 12807 |

| | | | |
|------------|---|--|-------|
| Qy | 2729 | ATCTCCAGACCGCAGCGCGCATCTCGGGCAGCGTTGGTCTCTGG- | 2776 |
| Db | 12808 | | 12867 |
| Qy | 2777 | ----- | 2776 |
| Db | 12868 | ATGATCGTCTCTCGTTGAGGACCGGCTAGGCTGGCGGGTTCCTTACTGTTAG | 12927 |
| Qy | 2777 | ----- | 2776 |
| Db | 12928 | CAGAAATGAATCACCGATACCGGACGGAACGTGAAGCGACTGCTGCTGCANAAACGTCTGCG | 12987 |
| Qy | 2777 | ----- | 2776 |
| Db | 12988 | ACCTGAGCAACAACATGAATGGTCTTCGGTTTCGGTTCGTAAGTCTGGAACGCGG | 13047 |
| Qy | 2777 | ----- | 2776 |
| Db | 13048 | AAGTCAGCGCCTGCACCATATATGTTCCGGATCTGCATCGCAGGATGCTGCTGCTACCC | 13107 |
| Qy | 2777 | -----GCTGGCATTTGACCTGAGTGATTTTCTC | 2805 |
| Db | 13108 | TGTGGAACACCTACATCTGTATTACGAAGCGCTGGCATTGACCTGAGTGATTTTCTC | 13167 |
| Qy | 2806 | TGTTCCCGCCGATCCATACCGCAGTGTTTTACCTCACACGTTCCAGTAACCGGCA | 2865 |
| Db | 13168 | | 13227 |
| Qy | 2866 | TGTTTCATCATCAGTAACCGCGTATCGTGAGCATCTCTCTCGTTTCATCGGTATCATTTACC | 2925 |
| Db | 13228 | TGTTTCATCATCAGTAACCGCGTATCGTGAGCATCTCTCTCGTTTCATCGGTATCATTTACC | 13287 |
| Qy | 2926 | CCCATGAACAGAAATCCCTTTACACGAGGATCAGTGACCAACAGAGAAAAACCGCC | 2985 |
| Db | 13288 | CCCATGAACAGAAATCCCTTTACACGAGGATCAGTGACCAACAGAGAAAAACCGCC | 13347 |
| Qy | 2986 | CTTAACATGCGCCGCTTTATCAGAGCCAGACATTAAGCTTCTGGAGAACTCAACGAG | 3045 |
| Db | 13348 | CTTAACATGCGCCGCTTTATCAGAGCCAGACATTAAGCTTCTGGAGAACTCAACGAG | 13407 |
| Qy | 3046 | CTGACGCGGATGAACAGCAGAGCTCT | 3073 |
| Db | 13408 | CTGACGCGGATGAACAGCAGACATCT | 13435 |
| RESULT 12 | AY648301 15270 bp DNA circular SYN 14-JAN-2005 | | |
| LOCUS | Dengue virus type 4 vector p4, complete sequence. | | |
| DEFINITION | AY648301 | | |
| ACCESSION | AY648301.1 GI:49781322 | | |
| VERSION | Dengue virus type 4 vector p4 | | |
| KEYWORDS | other sequences; artificial sequences; vectors. | | |
| SOURCE | 1 (bases 1 to 15270) | | |
| ORGANISM | Dengue virus type 4 vector p4 | | |
| REFERENCE | Dengue virus type 4 vector p4 | | |
| AUTHORS | other sequences; artificial sequences; vectors. | | |
| TITLE | 1 (bases 1 to 15270) | | |
| | Dengue virus type 4 vector p4 | | |
| JOURNAL | Dengue virus type 4 vector p4 | | |
| | other sequences; artificial sequences; vectors. | | |
| PUBMED | 1 (bases 1 to 15270) | | |
| REFERENCE | Dengue virus type 4 vector p4 | | |
| AUTHORS | other sequences; artificial sequences; vectors. | | |
| TITLE | 1 (bases 1 to 15270) | | |
| | Dengue virus type 4 vector p4 | | |
| JOURNAL | Dengue virus type 4 vector p4 | | |
| | other sequences; artificial sequences; vectors. | | |
| PUBMED | 1 (bases 1 to 15270) | | |
| REFERENCE | Dengue virus type 4 vector p4 | | |
| AUTHORS | other sequences; artificial sequences; vectors. | | |

| | |
|----------|---|
| TITLE | Direct Submission |
| JOURNAL | Submitted (08-JUN-2004) Laboratory of Infectious Diseases, NIAID, 50 South Drive, Bldg 50, Room 6515, Bethesda, MD 20892, USA |
| FEATURES | Location/Qualifiers |
| source | 1. 15270 |
| | /organism="Dengue virus type 4 vector p4" |
| | /mol_type="other DNA" |
| | /db_xref="taxon:283795" |
| | /focus |
| source | 1. 10649 |
| | /organism="Dengue virus type 4" |
| | /mol_type="other DNA" |
| | /db_xref="taxon:11070" |
| | /country="Dominica" |
| source | 10859. 15020 |
| | /organism="Cloning vector pBR322" |
| | /mol_type="other DNA" |
| | /db_xref="taxon:47470" |
| source | 15021. 15270 |
| | /organism="Cloning vector pGEM-3" |
| | /mol_type="other DNA" |
| | /db_xref="taxon:90108" |
| CDS | 102. 10265 |
| | /codon_start=1 |
| | /product="polyprotein precursor" |
| | /protein_id="AAT68472.1" |
| | /db_xref="GI:49781323" |
| | /translation="MNORKKVVPPFNMVKRNRVSTPQGLVKRFSGLSGKGLR |
| | MVLAFTPLRVLSIPTAGILRWGQKKNKAIKILIGFRKEIGRMILNMLNKRKSTI |
| | TLCLIPTVWAFSLSTRDEPLMTVAKHERGRPLLFKTEKINGTILAMLDGMCECD |
| | TVTVKPLLVNTEPEDIDWCNLTSTVMYGTCTQSGRERKRSVALTPHSGMGLT |
| | RAETWMSSEGAKHAQVESMILRNPGFALLAGPMAYMIGOTQRTVFFVLMMLVAP |
| | SYNRCVGVGNRDFEVSGGAWDLVLEHGGCVTTWAQKPTLDLFTLTKTAKVAL |
| | LRITCISASINIIITATRCPTQGPPELKEEQDQYICRRVDVDRGWGNGCGLFKGGV |
| | VTCAFSGSGKITCNLVENLEYTVVTVVTHNGDTSVHNGVTSVHNGVTSVHNGVTSV |
| | EVKLPDYCEITLDCPEPSGIDFENMILMKKKTKLWHKQFLLDPLPMTAGADTSEV |
| | HNKVRMTFKVPHAKRQDVTVLGSGQGAHMSALAGATEVDSGDMHMFAGHLCKV |
| | RMEKRLKMSYTWCSGKFSIDKEMAEOTHGCTTVVKVYEGAGAPCKVPIRIRDVKE |
| | KVGR1ISSPLAENTSNVNIIELEPPGDSYIVIGNSALTILHWPFGKSSIGWME |
| | STYGRKRAAILGETAMDVFGSGGLFTSLGKAVHQVFGSVTVTFGGVSWIRILIGF |
| | LVLWIGNSRNTSMATCIAVGGITLFLGFTVQADMGCVASWSKELCKGSGIFVVDN |
| | VHTWTEQVFKQFQSPARLASAILNAHKDGVGIRSTTRLENVMWKQITNELNYLWEG |
| | GHDLTVAGDVKGLTKGRALTPVSDLTQSWTKGAKIFTPPEARSTFLIDGPDT |
| | SECNERRANSLVEDYFGCMFTTNIMKREGSEVCDHRLMSAAIKDOKAHVADM |
| | GYWIESKNQWQIEKASLIEVKTLMPKHTLMSNGVLESQMLIPKSVAGPFPQHN |
| | RQYATQVTPGPHLGKLEIDFGCEPCTVTITQECDDHGRPSIRTTASGLVTPQCCR |
| | SCTMPPLRFLGEDCWTGMEIRPLSEKEENNVKQSVTAGQGTSETFFSGMLLCTLFVE |
| | ECLRRVTRKHMILVWITLCAILGLTMDLLRALIMLGDTSWGRIGGQTHLAIMA |
| | VFKMSPGVILGVFLKLTSRRTALMVIGMAMTIVLSIPHIDMELIDGSLGILIKIV |
| | TFDNTQVGTIALSLTFRSTMPVMAWRTIMAVLVVTLIPLCRTSCLOKQSHWEI |
| | TALLIGQAALPVYILMTMKGASRRSWPLNEGIMAVGLVSLGSLAKNDVPLAGPMVA |
| | GGLLLAAYVMSGADISLEKAAVQWMDADITGSSPIIEVQDEDSFSIRDEET |
| | NMITLVKALITVSGLYPLAIPVTMTLWYMWQVTKQSGALMDVPSPAATKAALSE |
| | GVYRIMORGLFKGTQVGVGHMEGVFHTMHVTRGVSICHTGRLEPESADVRNDMS |
| | YGGWRLLGDWKEEDVOVLAIEPKKPKHVQTPGLFKTLTGTEGTPKPTGTS |
| | SPILNRKGVILGXGVVTVKSGDVSAITQAEIGEPDYEVDIIPKRLTIDMLH |
| | PGAGTKRILPSIVREALKRRLRILILAPTVAAMEEALRGIPTQTPAVKSEHT |
| | GREIVDMCHATFTTFLSSSTKPNYINILVMDERHFTDPSSVAARGYISTRVMEGEA |
| | AIFMTATPGGATFPFQNSNIEDIERIPERSWNTGFWITDYGKTVFWFVPSIKAG |
| | NDIANCLRSKGVILSRKTFDTEYPKTLTDWDFVTTDISEMGANFRAGRLDPR |
| | RCLKPLIPDGPVILAGPIPVTPASAAQRGRIGRNPADDDQYVSGDGLKND |
| | HAHTEAKMLDNIYTPGIIPLFGPERKTOAIDGDFRLGRQKTFVLEMRGDL |
| | PWVLSYKVASAGISYKQRECMFTGERNNQIILEENNEVIRWREGKKRLPWLDAV |
| | YADPMALADPFKPSGRKSIITDILTEIASLPTVLSRAKALDNLVMTHTERGRA |
| | YOHANLPELSTELMLVALLGAMTAGIFLFGQKGIKLSMGLITITAVASGLLWVA |
| | BIQPMIAASIIIEFLFVLLIPEPEKQRTFDNQILYITLITILITGLIAANGLI |
| | EKTQTDGFGVTEITLIDVLRPASATLYAVATITLPMRLHRTIENTSANTSLAA |
| | INAAQAVLMGLKGMPLHRMDVPLLAMGVCSOVNPTTLTASLVLMVHYAIGPGL |
| | QAKATREAKRTAAGIMKNPTVDGITTVDLSPISYDPKFKOLGOVMLLVLCAGOLL |
| | MRTTAPCEVITLATGPIITLWEGNPGFWNTIIVASTANIFRGSLAGAGLAFSLIK |
| | NAQTPRGTTGTTGELGKWRKQLNSLDKGFEEYKRSKGLPMTKNTVEYTKGKGP |
| | KHVSRSKSKIRIVERGVNPKGVKVDLGGCGGWSYMATLKNVTEYTKGKGP |
| | HEEPIPMATYGNLWKLHSGVDVFKYKPTQEQVDTLLCDIGESSNPTIEBGRVRLVKM |


```
QY 1409 ACCGTGATGAATCTAAATGCGCTCATCGTCAATCTCGGCACCGTCAACCTGGATGC 1468
Db 11505 ACCGTGATGAATCTAAATGCGCTCATCGTCAATCTCGGCACCGTCAACCTGGATGC 11564
QY 1469 TGTAGGCATAGCTTTGGTTATGCGGTAATGCGGCTCTTGGGGATATCGTCCATTC 1528
Db 11565 TGTAGGCATAGCTTTGGTTATGCGGTAATGCGGCTCTTGGGGATATCGTCCATTC 11624
QY 1529 CGACAGCATGCCAGTCACTATGGCGTCTGCTAGCGCTATATGCGTTGATGCAATTTCT 1588
Db 11625 CGACAGCATGCCAGTCACTATGGCGTCTGCTAGCGCTATATGCGTTGATGCAATTTCT 11684
QY 1589 ATGCGCACCCGTTCTCGAGCACTGTCCGACGCTTTGGCGCGCCGCCAGTCTGCTGCG 1648
Db 11685 ATGCGCACCCGTTCTCGAGCACTGTCCGACGCTTTGGCGCGCCGCCAGTCTGCTGCG 11744
QY 1649 TTGCGTACTTGGAGCCACTATCGACTACGCGCATCATGCGGACCAACACCGCTCTGTGGAT 1708
Db 11745 TTGCGTACTTGGAGCCACTATCGACTACGCGCATCATGCGGACCAACACCGCTCTGTGGAT 11804
QY 1709 CCTCTACGCGAGCGCATCTGGCGCGGCATCACCGCGGCCACAGGTGCGGTTGCTGGCG 1768
Db 11805 CCTCTACGCGAGCGCATCTGGCGCGGCATCACCGCGGCCACAGGTGCGGTTGCTGGCG 11864
QY 1769 CTATATCGCGGACATCACCGATGGGGAAGATCGGGCTCGCCACTTTGGGGCTCATGAGCG 1828
Db 11865 CTATATCGCGGACATCACCGATGGGGAAGATCGGGCTCGCCACTTTGGGGCTCATGAGCG 11924
QY 1829 TTGTTTTCGGCGTGGTATGGTGGCAGGCCCGCTCGCGCGGGGAGTGTGGCGGCATCTC 1888
Db 11925 TTGTTTTCGGCGTGGTATGGTGGCAGGCCCGCTCGCGCGGGGAGTGTGGCGGCATCTC 11984
QY 1889 CTTGCATGCACTCTCTTGGCGCGGGTCTCTCAACCGGCCTCAACCTACTACTGCGGCTG 1948
Db 11985 CTTGCATGCACTCTCTTGGCGCGGGTCTCTCAACCGGCCTCAACCTACTACTGCGGCTG 12044
QY 1949 CTTCTTAATCAGAGTGCATTAAGGAGAGCGTGCAGCCGATGCTTGGAGAGCTTCAA 2008
Db 12045 CTTCTTAATCAGAGTGCATTAAGGAGAGCGTGCAGCCGATGCTTGGAGAGCTTCAA 12104
QY 2009 CCGAGTCAGCTCTTCGCTGGCGCGGGGATGACTATCGTCCGCGCATTTAGACTGT 2068
Db 12105 CCGAGTCAGCTCTTCGCTGGCGCGGGGATGACTATCGTCCGCGCATTTAGACTGT 12164
QY 2069 CTTCTTTATCATGCAACTCGTAGACAGGTGCGCGGAGCGCTCTGGGTCACTTTTCGGCG 2128
Db 12165 CTTCTTTATCATGCAACTCGTAGACAGGTGCGCGGAGCGCTCTGGGTCACTTTTCGGCG 12224
QY 2129 GGACCGCTTTCGCTGGAGCGCGATGATCGGCTGTGCTCGGTATTCGGAATCTT 2188
Db 12225 GGACCGCTTTCGCTGGAGCGCGATGATCGGCTGTGCTCGGTATTCGGAATCTT 12284
QY 2189 GCAGCGCTCGCTCAAGCTTCGTCAGTGTCCCGCCACCAAGCTTTCGCGGAGGCA 2248
Db 12285 GCAGCGCTCGCTCAAGCTTCGTCAGTGTCCCGCCACCAAGCTTTCGCGGAGGCA 12344
QY 2249 GGCCATTTATCGCGGATGGCGCGGACGCGCTGGGTACGTCTGTGCGGTTCGGGAC 2308
Db 12345 GGCCATTTATCGCGGATGGCGCGGACGCGCTGGGTACGTCTGTGCGGTTCGGGAC 12404
QY 2309 GCGAGGCTGGATGGCTTCCCAATTATGATTTCTCGCTTCGCGCGGCATCGGGATGCC 2368
Db 12405 GCGAGGCTGGATGGCTTCCCAATTATGATTTCTCGCTTCGCGCGGCATCGGGATGCC 12464
QY 2369 CGGTTGAGGCGCATGCTGTCAGGCGAGGTAGATGACGACCATCAGGACGCTTCAAG 2428
Db 12465 CGGTTGAGGCGCATGCTGTCAGGCGAGGTAGATGACGACCATCAGGACGCTTCAAG 12524
QY 2429 ATCGCTCGGCTCTTACGAGCCTAACTTCGATCATTTGACCGCTGATCGTCACGGCGAT 2488
Db 12525 ATCGCTCGGCTCTTACGAGCCTAACTTCGATCATTTGACCGCTGATCGTCACGGCGAT 12584
```

```
QY 2489 TTATGCCCTCGCGGAGCACATGAAACGGGTTGGCATGGATTGTAGGCGCGCCCTATA 2548
Db 12585 TTATGCCCTCGCGGAGCACATGAAACGGGTTGGCATGGATTGTAGGCGCGCCCTATA 12644
QY 2549 CTTGTCTGCTCTCCCGGTTGCGGTGCGGTGCAATGAGCGCGGCCACTCGACCTGAAT 2608
Db 12645 CTTGTCTGCTCTCCCGGTTGCGGTGCGGTGCAATGAGCGCGGCCACTCGACCTGAAT 12704
QY 2609 GGAAGCGCGGCGCACTCGCTAAACGGATTCAACCAATTCGAAGATTTGGAGCCAATCAATTC 2668
Db 12705 GGAAGCGCGGCGCACTCGCTAAACGGATTCAACCAATTCGAAGATTTGGAGCCAATCAATTC 12764
QY 2669 TTGCGGAGAACTGTAATCGCAAAACCAACCTTTGGGAGAACATATCCATGCGTCCGCC 2728
Db 12765 TTGCGGAGAACTGTAATCGCAAAACCAACCTTTGGGAGAACATATCCATGCGTCCGCC 12824
QY 2729 ATCTCCAGCAGCCGCGCGCATCTCGGGGAGCGTTGGGTCTCTGG----- 2776
Db 12825 ATCTCCAGCAGCCGCGCGCATCTCGGGGAGCGTTGGGTCTCTGGCCACGGGTGCGC 12884
QY 2777 ----- 2776
Db 12885 ATGATGCTGCTCTGTCGTTGAGGACCGCGCTAGGCTGGCGGGTTGCTTACTGCTTAG 12944
QY 2777 ----- 2776
Db 12945 CAGAATGAATCACCGGATACGCGAGCGAAACGTTGAAGCGACTGCTGCTCAAAACGCTCGC 13004
QY 2777 ----- 2776
Db 13005 ACCTGAGCAACACATGAATGGTCTTCGTTTCGTTTCGTTAAAGTCTCGAAACGCGG 13064
QY 2777 ----- 2776
Db 13065 AAGTCAGCGCCCTGCACCATTAATGTTCCGGATCTGCATCGCAGGATGCTGCTGGCTACCC 13124
QY 2777 -----GTTGGCATTCGACCTCGAGTGATTTTCTC 2805
Db 13125 TGTGAAACACCTACATCTGATTAACGAAGCGCTGGCATTCACCTGAGTGATTTTCTC 13184
QY 2806 TGGTCCCGCGCATCCATACCGCAGTTGTTTACCTCACACGTTCCAGTAACCGGGCA 2865
Db 13185 TGGTCCCGCGCATCCATACCGCAGTTGTTTACCTCACACGTTCCAGTAACCGGGCA 13244
QY 2866 TGTTTCATCATCAGTAACCCGCTATCGTAGCATCTCTCTCGTTTCATCGGTATCATTTACC 2925
Db 13245 TGTTTCATCATCAGTAACCCGCTATCGTAGCATCTCTCTCGTTTCATCGGTATCATTTACC 13304
QY 2926 CCCATGAACAGAAATCCCTTTACACGAGGAGCATCAGTGACCAACAGGAAAAAACCGCC 2985
Db 13305 CCCATGAACAGAAATCCCTTTACACGAGGAGCATCAGTGACCAACAGGAAAAAACCGCC 13364
QY 2986 CTTAACATGGCCGCTTTATCAGAAGCCAGACATTAACGTTCTGGAGAAACTCAACGAG 3045
Db 13365 CTTAACATGGCCGCTTTATCAGAAGCCAGACATTAACGTTCTGGAGAAACTCAACGAG 13424
QY 3046 CTGGACCGGGATGAACAGGCGAGCTCT 3073
Db 13425 CTGGACCGGGATGAACAGGCGAGCATCT 13452
```

RESULT 14

AY028776/c

LOCUS

AY028776

DEFINITION

AY028776

VERSION

AY028776.1

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

13561 bp DNA circular SYN 22-DEC-2004
TphoZ mutagenesis vector pMH120, complete sequence.

GI:13549398

TphoZ mutagenesis vector pMH120
TphoZ mutagenesis vector pMH120

other sequences; artificial sequences; vectors.
1 (bases 1 to 13561)

Lee, M.H., Nittayajarn, A., Ross, R.P., Rothschild, C.B., Parsonage, D.,

| | | | |
|-----------|--|---------------------|--------|
| TITLE | Claiborne, A. and Rubens, C.E. | | |
| | Characterization of <i>Enterococcus faecalis</i> alkaline phosphatase and use in identifying <i>Streptococcus agalactiae</i> secreted proteins | | |
| JOURNAL | J. Bacteriol. | 181 (18), 5790-5799 | (1999) |
| PUBMED | 10482522 | | |
| REFERENCE | 2 | (bases 1 to 13561) | |
| AUTHORS | Clancy, A., Lee, M.H., Jones, A.L. and Rubens, C.E. | | |
| TITLE | Construction and characterization of transposon TnpHoz for the identification of genes encoding exported proteins in <i>Streptococcus agalactiae</i> | | |
| | FEMS Microbiol. Lett. 241 (2), 257-264 (2004) | | |
| JOURNAL | 15598541 | | |
| PUBMED | 3 | (bases 1 to 13561) | |
| REFERENCE | Lee, M.H., Clancy, A., Jones, A.L., Nittayajarn, A. and Rubens, C.E. | | |
| AUTHORS | Direct Submission | | |
| TITLE | Submitted (19-MAR-2001) Infectious Disease, Children's Hospital and Regional Medical Center, 4800 Sand Point Way NE, Seattle, WA 98105, USA | | |
| | USA | | |
| FEATURES | Location/Qualifiers | | |
| | 1..13561 | | |
| source | /organism="TnpHoz mutagenesis vector pMH120" | | |
| | /mol_type="genomic DNA" | | |
| gene | /db_xref="taxon:156762" | | |
| | /lab_host="Streptococcus agalactiae" | | |
| CDS | /note="transposon TnpHoz delivery vehicle" | | |
| | complement(4167..4928) | | |
| gene | /gene="erm" | | |
| | complement(4167..4928) | | |
| CDS | /gene="erm" | | |
| | /note="confers erythromycin-resistance" | | |
| gene | /codon_start=1 | | |
| | /transl_table=11 | | |
| CDS | /product="rRNA methylase" | | |
| | /protein_id="AAK27828.1" | | |
| gene | /db_xref="GI:13549399" | | |
| | /translation="MYHFRSDYMMNKNIKYSONFLTNBKNRNLQIILKLNKEDTDTVYBI GTGGHLLTKLAKTSKOVTSYELDSHLFNLSEKLNIRVTLHODILOFQFPNKR YKIVGNIPHLSTOILIKVWPESHASDIYLIYVEGFYKRTLDIHRITGLLLHTOVSIQ QLLKPAECFHPKPKNSVLKILTRHTDVEDPKWKLTYTFVSKWNRVRLFTKQ FQAMKHAKNVNLSTVYEQVLSIFNSLLEFNGRK" | | |
| CDS | complement(6372..6563) | | |
| | /gene="rop" | | |
| gene | complement(6372..6563) | | |
| | /gene="rop" | | |
| CDS | /note="regulatory protein; involved in determination of plasmid copy number" | | |
| | /codon_start=1 | | |
| gene | /transl_table=11 | | |
| | /product="ROP protein" | | |
| CDS | /protein_id="AAK27829.1" | | |
| | /db_xref="GI:13549400" | | |
| gene | /translation="MTQDEKFTALNWARPIRSQTLTLEKLNELDAEQADICESLHDH ADELYRSLARFGDGENL" | | |
| | complement(7201..8391) | | |
| CDS | /gene="tet" | | |
| | complement(7201..8391) | | |
| gene | /gene="tet" | | |
| | /codon_start=1 | | |
| CDS | /transl_table=11 | | |
| | /product="tetracycline-resistance protein" | | |
| gene | /protein_id="AAK27830.1" | | |
| | /db_xref="GI:13549401" | | |
| CDS | /translation="MKSNNALIVILGTVTLDAVGILVWPVLPGLLRDIIVHSDSIASH YGVLLALYMQFLCAPVILGALSIRGFRPVLASLUGATIDYAIMATTIPVIMILYAG RIVAGITGAVAGAYTADITDGEDRARHFLMSACFVGWVAGPVGGLGALSILH APFLAAVANGNLGILGFLMQESHKGERRPMPAFNPVSVFRWAGMTIYAALMTV FFIQLVQVPAALVWIFGEDRFRWSATMI GLSLAVFGILHALAQAFVTGPATKRFGE KOALTAGNAWALVIFATRGWMAFPIMLLASGIGMPALQAMLSROVDDHOG QUGSLAALTSLTSITGFLI VTAIYAASASTWNGLAVIGVLAALVIVCLPALRRGWSR ATST" | | |
| | complement(8683..10032) | | |
| gene | /gene="phoZ" | | |
| | /allele="phoZ2" | | |

| | | | |
|----------------------------|--|---|------|
| gene | /notes="alkaline phosphatase; derived from <i>Enterococcus faecalis</i> ; expression is dependent upon integration event to create in-frame start codon" | | |
| | 10619..11413 | | |
| CDS | /gene="aphA3" | | |
| | 10619..11413 | | |
| gene | /gene="aphA3" | | |
| | /notes="confers kanamycin-resistance" | | |
| CDS | /codon_start=1 | | |
| | /transl_table=11 | | |
| gene | /product="aminoglycoside phosphotransferase A3" | | |
| | /protein_id="AAK27831.1" | | |
| CDS | /db_xref="GI:13549402" | | |
| | /translation="MAKWRISPELKKLIEKYRCVKDTGMSPAKYKLVGENENLYLK MTSRYKGTVDYVEREKDMLLEGLVPKVLHFERHDGWSNLMSEADGVLCSSEY EDQSPKIIELAECLRLFHSIDISDPCVTNSLDSRLAELDYLLNNDLADVDCCENWE EDTFPKDRIFDLTKERPEBELVPSHGLDGSNIFVKDGKVSFGFIDLGRSGRADKW YDTAFVRSIREDIGEQYVELFDLLGLIKPDWEKIKIYYILLDLF" | | |
| gene | complement(11817..12515) | | |
| | /gene="repA3" | | |
| CDS | complement(11817..12515) | | |
| | /gene="repA3" | | |
| gene | /codon_start=1 | | |
| | /transl_table=11 | | |
| CDS | /product="temperature-sensitive RepA" | | |
| | /protein_id="AAK27832.1" | | |
| gene | /db_xref="GI:13549403" | | |
| | /translation="MAIKNTKARNFGLLYPDSIPNDWKEKLSIGVSMVSPHLMD EKOKDTWNNSNIQNGKHKKHYHVIYIARPVITLESVRNKKRLGNSSVAHVLEI LDYKGSYELTHESKDAIKNKHIDKOLINDFIDRIYITLDSQKRELKNLL LDIVDYNLVNLTQDLMAFIRLGAEFGLINTNDVKDITVSTSSAFRLFEGNYQCGYRA SYAKVLDAETGEIK" | | |
| Query Match | 34.0%; Score 1635.8; DB 11; Length 13561; | | |
| | Best Local Similarity 99.9%; Pred. No. 0; | | |
| Matches 1637; Conservative | 0; Mismatches | | |
| | 2; Indels | | |
| Gaps | 0; | | |
| | | | |
| Qy | 1140 | ATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTTATTT | 1199 |
| Db | 8667 | ATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTTATTT | 8608 |
| Qy | 1200 | AGAAAAATAACAAATAGGGTTCGCGCACATTTCCCGGAAAAAGTGCCACTGACGTCT | 1259 |
| Db | 8607 | AGAAAAATAACAAATAGGGTTCGCGCACATTTCCCGGAAAAAGTGCCACTGACGTCT | 8548 |
| Qy | 1260 | AAGAAACCATTTATCATGACATTAACCTATAAAATAGCGGTATCAGGAGCCCTTTC | 1319 |
| Db | 8547 | AAGAAACCATTTATCATGACATTAACCTATAAAATAGCGGTATCAGGAGCCCTTTC | 8488 |
| Qy | 1320 | GTCTTCAAGATTCTCATGTTTGAAGCTTATCATCGATAAGCTTTAATGCGGTAGTTTA | 1379 |
| Db | 8487 | GTCTTCAAGATTCTCATGTTTGAAGCTTATCATCGATAAGCTTTAATGCGGTAGTTTA | 8428 |
| Qy | 1380 | TCACAGTTAAATTCCTAACCGAGTCAGSCACCGTGTATGAAATCTAACAAATGCGGTATC | 1439 |
| Db | 8427 | TCACAGTTAAATTCCTAACCGAGTCAGSCACCGTGTATGAAATCTAACAAATGCGGTATC | 8368 |
| Qy | 1440 | GTCTTCTCGGACCGTCACCCCTGGATCTGTAGGCATAGGCTTGGTTATGCCGCTACTG | 1499 |
| Db | 8367 | GTCTTCTCGGACCGTCACCCCTGGATCTGTAGGCATAGGCTTGGTTATGCCGCTACTG | 8308 |
| Qy | 1500 | CGGGCCCTTTCGGGATATCGTCATTCCGACAGCATCGGCAGTCACCTATGGCGCTGCTG | 1559 |
| Db | 8307 | CGGGCCCTTTCGGGATATCGTCATTCCGACAGCATCGGCAGTCACCTATGGCGCTGCTG | 8248 |
| Qy | 1560 | CTAGCGCTATATGCGTTGATCAATTTCTATGCGCACCCGTTCTCGGAGCACTGTCCGAC | 1619 |
| Db | 8247 | CTAGCGCTATATGCGTTGATCAATTTCTATGCGCACCCGTTCTCGGAGCACTGTCCGAC | 8188 |
| Qy | 1620 | CGCTTTGGCCGCGCCGAGTCTGCTCGCTTTCGCTACTTGGAGCCACTATCGACTACCGG | 1679 |
| Db | 8187 | CGCTTTGGCCGCGCCGAGTCTGCTCGCTTTCGCTACTTGGAGCCACTATCGACTACCGG | 8128 |

ORIGIN

Qy 1680 ATCATGGCGACACACCCGCTCTGTGGATCTCTACGCCGAGCAGCATCGTGGCCGGCAtC 1739
Db 8127 ATCATGGCGACACACCCGCTCTGTGGATCTCTACGCCGAGCAGCATCGTGGCCGGCAtC 8068
Qy 1740 ACCGCGCCACAGGTGGCGCTTCTGGCGCTATATCGCCGACATCACCGATGGGGAAGAT 1799
Db 8067 ACCGCGCCACAGGTGGCGCTTCTGGCGCTATATCGCCGACATCACCGATGGGGAAGAT 8008
Qy 1800 CGGGCTCGCCACTTCGGGCTCATGAGCGCTTGTGTTTCGGCGTGGGTATGTTGGCAGCGCCC 1859
Db 8007 CGGGCTCGCCACTTCGGGCTCATGAGCGCTTGTGTTTCGGCGTGGGTATGTTGGCAGCGCCC 7948
Qy 1860 GTGGCCGGGAGACTGTTGGGCGCATCTCTTTCGATGACACATTCCTTGGCGCGCGGTG 1919
Db 7947 GTGGCCGGGAGACTGTTGGGCGCATCTCTTTCGATGACACATTCCTTGGCGCGCGGTG 7888
Qy 1920 CTCAAGCGCTCAACCTACTACTGCGCTTCTTAATGCGAGGAGTGCATAAAGGAGAG 1979
Db 7887 CTCAAGCGCTCAACCTACTACTGCGCTTCTTAATGCGAGGAGTGCATAAAGGAGAG 7828
Qy 1980 CGTGACCGATGCCCTTGAGAGCCTTCAACCCAGTCAGCTCTTCGGTGGCGCGGGCC 2039
Db 7827 CGTGACCGATGCCCTTGAGAGCCTTCAACCCAGTCAGCTCTTCGGTGGCGCGGGCC 7768
Qy 2040 ATGACTATCGTGGCGCACTTATGACTCTCTTCTTATCATGCAACTCGTAGGACAGGTG 2099
Db 7767 ATGACTATCGTGGCGCACTTATGACTCTCTTCTTATCATGCAACTCGTAGGACAGGTG 7708
Qy 2100 CCGCGAGCGCTCTGGGTCAATTTTCGGCGAGGACCGCTTTCCTGGAGCGCCAGCATGATC 2159
Db 7707 CCGCGAGCGCTCTGGGTCAATTTTCGGCGAGGACCGCTTTCCTGGAGCGCCAGCATGATC 7648
Qy 2160 GGCCTGTGCTTGGGTATTCGGAATCTTGCAAGCGCTTCGCTCAAGCGCTTCTGCTCACTGGT 2219
Db 7647 GGCCTGTGCTTGGGTATTCGGAATCTTGCAAGCGCTTCGCTCAAGCGCTTCTGCTCACTGGT 7588
Qy 2220 CCGCGCAACAAAGTTTCGGCGAGAGCAGGCAATTTATCGCGGAGCGCGCGCGAGCG 2279
Db 7587 CCGCGCAACAAAGTTTCGGCGAGAGCAGGCAATTTATCGCGGAGCGCGCGCGAGCG 7528
Qy 2280 CTGGGCTACGCTTCTGGGCTTCGCGAGCGAGGCTGATGGCTTCCCATATGAT 2339
Db 7527 CTGGGCTACGCTTCTGGGCTTCGCGAGCGAGGCTGATGGCTTCCCATATGAT 7468
Qy 2340 CTCTCTGCTTCGGCGGATCGGATGCCGCTTGCGAGGCGATGCTGTCAGGCAAGT 2399
Db 7467 CTCTCTGCTTCGGCGGATCGGATGCCGCTTGCGAGGCGATGCTGTCAGGCAAGT 7408
Qy 2400 GATGACGACCATCAGGAGCAGCTTCAAGGATCGCTCGCGGCTTCTTACGAGCTTAACCTG 2459
Db 7407 GATGACGACCATCAGGAGCAGCTTCAAGGATCGCTCGCGGCTTCTTACGAGCTTAACCTG 7348
Qy 2460 ATCATGACCGCTGATCGTCACGGCGATTTATGCGCGCTCGGCGAGCAGATGGAACGG 2519
Db 7347 ATCATGACCGCTGATCGTCACGGCGATTTATGCGCGCTCGGCGAGCAGATGGAACGG 7288
Qy 2520 TTGCGATGGATTGAGGCGCGCCTATACCTTGTCTGCGCTCCCGCGCTTCTGCGCGGT 2579
Db 7287 TTGCGATGGATTGAGGCGCGCCTATACCTTGTCTGCGCTCCCGCGCTTCTGCGCGGT 7228
Qy 2580 GCATGGAGCGCGGCGCACCTCGACCTGGAATGGAAGCGCGCGGCACTCGCTTAACGGATTCA 2639
Db 7227 GCATGGAGCGCGGCGCACCTCGACCTGGAATGGAAGCGCGCGGCACTCGCTTAACGGATTCA 7168
Qy 2640 CCATCCAGAAATGGAGCCAAATCAATTTCTTGGGAGAACTGTGAATGCGCAAAACCAACC 2699
Db 7167 CCATCCAGAAATGGAGCCAAATCAATTTCTTGGGAGAACTGTGAATGCGCAAAACCAACC 7108
Qy 2700 CTTGCGAGAACATATCCATCGCTCGCGCATCTCCAGCAGCGCGCGGCGCATCTCGG 2759
Db 7107 CTTGCGAGAACATATCCATCGCTCGCGCATCTCCAGCAGCGCGCGGCGCATCTCGG 7048

Qy 2760 GCAGGTTGGGTCTCTGGGC 2778
Db 7047 GCAGGTTGGGTCTCTGGGC 7029
RESULT 15
SYNCLCTAPS
LOCUS
DEFINITION
Expression cloning vector pNP6CRMCTAP colicin E1-connective tissue
activating peptide fusion protein gene, complete cds.
ACCESSION
M88538
VERSION
M88538.1 GI:208128
KEYWORDS
colicin E1; colicin:E1-CTAP-III(Leu21) fusion protein; connective
tissue activating peptide-III; expression vector.
SOURCE
unidentified cloning vector
ORGANISM
other sequences; artificial sequences; vectors.
REFERENCE
1 (bases 1 to 6343)
Waleh,N.S., Sohel,I., Lazar,J.B., Hudson,D.V., Sze,P.,
Underhill,P.A. and Johnson,P.H.
TITLE
Expression of synthetic genes encoding fused proteins under tight
control of modified regulatory regions of the colicin operon
JOURNAL
Gene 117 (1), 7-14 (1992)
PUBMED
1644316
COMMENT
Original source text: Cloning vector DNA.
FEATURES
Location/Qualifiers
1..6343
/organism="unidentified cloning vector"
/mol_type="genomic DNA"
/db_xref="taxon:45196"
-35_signal
5794..5799
-10_signal
5817..5822
promoter
5829..5861
RBS
5895..5899
CDS
5907..6182
/evidence=experimental
/codon_start=1
/transl_table=11
/product="Col-CTAP-III(Leu21) small fusion protein"
/protein_id="AA73219.1"
/db_xref="GI:208129"
/translation="METAVMNLAKBESLSDLYAELRCLIKTTSIHFKNIQSL
VIGKTHCNQVEVIATUDGRKICLDPAPIKKIVQKLAGDSAD"
sig_peptide
5907..5924
/note="leader sequence; putative"
mat_peptide
5925..6179
/product="connective tissue activating peptide III"
/standard_name="CTAP-III"
/note="(Leu21); CTAP-III(Leu21); putative"
terminator
6278..6308
/standard_name="Rho-dependent terminator"
ORIGIN
Query Match 33.7%; Score 1624; DB 11; Length 6343;
Best Local Similarity 98.6%; Pred. No. 0;
Matches 1650; Conservative 0; Mismatches 20; Indels 4; Gaps 1;
Qy 1109 TGTGGAACACCTACATCTGTATTAAACGAAGCATTTATGAAGCATTTATCAGGTTATTGT 1168
Db 529 TGTGGAATCTCATCTCTCTTTTCAATATTATTGAAGCATTTATCAGGTTATTGT 588
Qy 1169 CTCATGAGCGGATACATATTTGAATGTATTAGAAAAATAAACAATAGGGGTTCCGGGC 1228
Db 589 CTCATGAGCGGATACATATTTGAATGTATTAGAAAAATAAACAATAGGGGTTCCGGGC 648
Qy 1229 ACATTTCCCGAAAGTGCACCTGACGCTTAAGAACCATTTATCATGACATTAAACC 1288
Db 649 ACATTTCCCGAAAGTGCACCTGACGCTTAAGAACCATTTATCATGACATTAAACC 708
Qy 1289 TATAAAAATAGCGCTATCACGAGCGCTTTCGTTCTTCAAG---AATTCATGTTTTCAC 1344
Db 709 TATAAAAATAGCGCTATCACGAGCGCTTTCGTTCTTCAAGAAATTAATTCATGTTTTCAC 768

| | | | |
|----|------|--|------|
| Qy | 1345 | AGCTTATCATCGATAAGCTTTAAATGCGGTAGTTTATACAGTAAATGCTAAACGAGTC | 1404 |
| Db | 769 | AGCTTATCATCGATAAGCTTTAAATGCGGTAGTTTATACAGTAAATGCTAAACGAGTC | 828 |
| Qy | 1405 | AGGCACCGTGATGAAATCTAAACATGCGCTCATCGTCATCTCGGCACCGTACCCCTGG | 1464 |
| Db | 829 | AGGCACCGTGATGAAATCTAAACATGCGCTCATCGTCATCTCGGCACCGTACCCCTGG | 888 |
| Qy | 1465 | ATGCTGTAGGATAGGCTTGTTATGCGGTACTGCGGGCTCTTGGGGATATCGTCC | 1524 |
| Db | 889 | ATGCTGTAGGATAGGCTTGTTATGCGGTACTGCGGGCTCTTGGGGATATCGTCC | 948 |
| Qy | 1525 | ATTCCGACAGCATCGCAGTCACTATGSCGTGCTAGCGCTATATGCGTTGATGAAT | 1584 |
| Db | 949 | ATTCCGACAGCATCGCAGTCACTATGSCGTGCTAGCGCTATATGCGTTGATGAAT | 1008 |
| Qy | 1585 | TTCTATGCGCACCGGTTCTCGGAGCACTGTCCGACCGCTTTGGCGCGGCCAGTCTCTGC | 1644 |
| Db | 1009 | TTCTATGCGCACCGGTTCTCGGAGCACTGTCCGACCGCTTTGGCGCGGCCAGTCTCTGC | 1068 |
| Qy | 1645 | TCGCTTCGCTACTTTGGAGCCACTATCGACTACGGGATCATGGCACACCCGTCCTGT | 1704 |
| Db | 1069 | TCGCTTCGCTACTTTGGAGCCACTATCGACTACGGGATCATGGGACACACCCGTCCTGT | 1128 |
| Qy | 1705 | GGATCCTCTACGCGGAGCATGCTGGCGGCGCATCACCGGCGCCACAGGTGCGGTGCTG | 1764 |
| Db | 1129 | GGATCCTCTACGCGGAGCATGCTGGCGGCGCATCACCGGCGCCACAGGTGCGGTGCTG | 1188 |
| Qy | 1765 | CGCGCTATATCGCGACATCACCGATGGGGAAGATCGGGCTCGCCACTTCGGGGCTCATGA | 1824 |
| Db | 1189 | CGCGCTATATCGCGACATCACCGATGGGGAAGATCGGGCTCGCCACTTCGGGGCTCATGA | 1248 |
| Qy | 1825 | CGCTGTGTTTCGGCGTGCGTATGCTGGCAGCGCCGCTGGCGGGGACTGTTGGCGCGCA | 1884 |
| Db | 1249 | CGCTGTGTTTCGGCGTGCGTATGCTGGCAGCGCCGCTGGCGGGGACTGTTGGCGCGCA | 1308 |
| Qy | 1885 | TCTCCTTCATGCACTACCACTTCCTGGCGGCGGTGCTCAACGGCTCAACCTACTCTG | 1944 |
| Db | 1309 | TCTCCTTCATGCACTACCACTTCCTGGCGGCGGTGCTCAACGGCTCAACCTACTCTG | 1368 |
| Qy | 1945 | GCTGCTTCTTAATGAGGAGTTCGATAGGAGAGCGTCGACCGATGCCCTTGAGAGCT | 2004 |
| Db | 1369 | GCTGCTTCTTAATGAGGAGTTCGATAGGAGAGCGTCGACCGATGCCCTTGAGAGCT | 1428 |
| Qy | 2005 | TCAACCCAGTCAGCTTCCTCGGTGGCGCGGGGATGACTATCGTCCCGCACTTATGA | 2064 |
| Db | 1429 | TCAACCCAGTCAGCTTCCTCGGTGGCGCGGGGATGACTATCGTCCCGCACTTATGA | 1488 |
| Qy | 2065 | CTGCTCTTCTTATCATGCAACTCGTAGGACAGGTGCGGCGCGCTCTGGGTCAATTTTCG | 2124 |
| Db | 1489 | CTGCTCTTCTTATCATGCAACTCGTAGGACAGGTGCGGCGCGCTCTGGGTCAATTTTCG | 1548 |
| Qy | 2125 | GCGAGGACCGCTTTCGCTGGAGCGCGAGATGATCGGCTGTGCTTGGGTATTCGGAA | 2184 |
| Db | 1549 | GCGAGGACCGCTTTCGCTGGAGCGCGAGATGATCGGCTGTGCTTGGGTATTCGGAA | 1608 |
| Qy | 2185 | TCTTGCA CGCCCTCGCTCAAGCTTTCGTCATGCTGTCGCCCAACAAAGTTTCGGCGAGA | 2244 |
| Db | 1609 | TCTTGCA CGCCCTCGCTCAAGCTTTCGTCATGCTGTCGCCCAACAAAGTTTCGGCGAGA | 1668 |
| Qy | 2245 | AGCAGGCCATTTATCGCGGATGGCGGCGGAGCGCTGGGCTACGCTTGTGGGGTTTCG | 2304 |
| Db | 1669 | AGCAGGCCATTTATCGCGGATGGCGGCGGAGCGCTGGGCTACGCTTGTGGGGTTTCG | 1728 |
| Qy | 2305 | CGACCGGAGGCTTGGCTTCCCACTTATGATTTCTCGCTTCCGCGGCGCATCGGGA | 2364 |
| Db | 1729 | CGACCGGAGGCTTGGATGGCTTCCCACTTATGATTTCTCGCTTCCGCGGCGCATCGGGA | 1788 |
| Qy | 2365 | TGCCCGGTTGAGGCCATGCTGTCAGGAGGTAGATGACGACCATCAGGGACAGCTTC | 2424 |
| Db | 1789 | TGCCCGGTTGAGGCCATGCTGTCAGGAGGTAGATGACGACCATCAGGGACAGCTTC | 1848 |
| Qy | 2425 | AAGGATCGCTCGCGGCTCTTACCAGCCTAACTTTCGATCATTTGGACCGCTGATCGTCACGG | 2484 |

Search completed: January 18, 2006, 04:26:11
Job time : 24858 secs

| | | | |
|----|------|--|------|
| Db | 1849 | AAGGATCGCTCGCGGCTCTTTACCGAGCCTAACTTCGATCACTGGACCGCTGATCGTCACGG | 1908 |
| Qy | 2485 | CGATTTATGCGGCTCGGCGAGCACATGGAAACGGGTTGGCATGGATTGTAGGCGCGCC | 2544 |
| Db | 1909 | CGATTTATGCGGCTCGGCGAGCACATGGAAACGGGTTGGCATGGATTGTAGGCGCGCC | 1968 |
| Qy | 2545 | TATACCTTGTCTGCTCCCGGCTTGGCTCGCGGTGATGGAGCCGGGCCACCTCGACCT | 2604 |
| Db | 1969 | TATACCTTGTCTGCTCCCGGCTTGGCTCGCGGTGATGGAGCCGGGCCACCTCGACCT | 2028 |
| Qy | 2605 | GAATGGAAGCGCGGCGCACCTCGCTAACGGAATTCACCTCCAAAGAAATTGGAGCAAATCA | 2664 |
| Db | 2029 | GAATGGAAGCGCGGCGCACCTCGCTAACGGAATTCACCTCCAAAGAAATTGGAGCAAATCA | 2088 |
| Qy | 2665 | ATTCTTGGGAGAACTGTGAATGGGCAACCAACCCCTTGGCAGAACATATCCATCGGTC | 2724 |
| Db | 2089 | ATTCTTGGGAGAACTGTGAATGGGCAACCAACCCCTTGGCAGAACATATCCATCGGTC | 2148 |
| Qy | 2725 | CGCCATCTCCAGCAGCGCGCACCGGCGCATCTCGGGCAGCGTTCGGTTCCTGGGC | 2778 |
| Db | 2149 | CGCCATCTCCAGCAGCGCGCACCGGCGCATCTCGGGCAGCGTTCGGTTCCTGGGC | 2202 |

GenCore version 5.1.6
Copyright (c) 1993 - 2006 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 17, 2006, 17:02:23 ; Search time 2993.54 Seconds
(without alignments)
10717.687 Million cell updates/sec

Title: US-10-511-327-7
Perfect score: 4814
Sequence: 1 ggtacctgtgaatgcgca.....taagatgctttctgtgact 4814

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_21:.*
1: Geneseqn1980s:.*
2: Geneseqn1990s:.*
3: Geneseqn2000s:.*
4: Geneseqn2001as:.*
5: Geneseqn2001bs:.*
6: Geneseqn2002as:.*
7: Geneseqn2002bs:.*
8: Geneseqn2003as:.*
9: Geneseqn2003bs:.*
10: Geneseqn2003cs:.*
11: Geneseqn2003ds:.*
12: Geneseqn2004as:.*
13: Geneseqn2004bs:.*
14: Geneseqn2005s:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|-------|-------------|
| 1 | 4814 | 100.0 | 4814 | 10 | ACF58174 |
| 2 | 1636.8 | 34.0 | 15153 | 12 | ADG33315 |
| 3 | 1636.8 | 34.0 | 15159 | 12 | ADG33313 |
| 4 | 1560.6 | 32.4 | 7311 | 14 | AEA06371 |
| 5 | 1477.2 | 30.7 | 5926 | 2 | AAV32978 |
| 6 | 1477.2 | 30.7 | 5926 | 6 | AAV32978 |
| 7 | 1476 | 30.7 | 5926 | 2 | AAV32977 |
| 8 | 1476 | 30.7 | 5926 | 6 | AAV32977 |
| 9 | 1476 | 30.7 | 5926 | 12 | ADG46817 |
| 10 | 1474 | 30.6 | 5926 | 12 | ADG46818 |
| 11 | 1447.8 | 30.1 | 4840 | 4 | AAV59457 |
| 12 | 1446.8 | 30.1 | 3474 | 2 | AAQ29146 |
| 13 | 1446.8 | 30.1 | 3474 | 2 | AAQ51546 |
| 14 | 1446.8 | 30.1 | 3801 | 12 | ADL72227 |
| 15 | 1446.8 | 30.1 | 4245 | 8 | ABZ70623 |
| 16 | 1446.8 | 30.1 | 5201 | 12 | ADL72228 |
| 17 | 1446.8 | 30.1 | 11184 | 12 | ADP86274 |
| 18 | 1446.8 | 30.1 | 11184 | 12 | ADP86276 |
| 19 | 1446.8 | 30.1 | 11184 | 12 | ADP86277 |

| | | | | | |
|----|--------|------|-------|----|----------|
| 20 | 1446.8 | 30.1 | 11313 | 12 | ADP86273 |
| 21 | 1446.8 | 30.1 | 11313 | 12 | ADP86264 |
| 22 | 1446.8 | 30.1 | 11313 | 12 | ADP86266 |
| 23 | 1446.8 | 30.1 | 11313 | 12 | ADP86265 |
| 24 | 1446.8 | 30.1 | 11313 | 12 | ADP86268 |
| 25 | 1446.8 | 30.1 | 11313 | 12 | ADP86270 |
| 26 | 1446.8 | 30.1 | 11313 | 12 | ADP86271 |
| 27 | 1446.8 | 30.1 | 11313 | 12 | ADP86272 |
| 28 | 1446.8 | 30.1 | 11313 | 12 | ADP86269 |
| 29 | 1446.8 | 30.1 | 11313 | 12 | ADP86275 |
| 30 | 1446.8 | 30.1 | 11313 | 12 | ADP86267 |
| 31 | 1446.8 | 30.1 | 12305 | 13 | ADT77488 |
| 32 | 1446.8 | 30.1 | 12306 | 10 | ADI41414 |
| 33 | 1446.8 | 30.1 | 12311 | 13 | ADT77500 |
| 34 | 1446.8 | 30.1 | 12315 | 10 | ADI41413 |
| 35 | 1446.8 | 30.1 | 12980 | 6 | AAV59364 |
| 36 | 1446.8 | 30.1 | 12980 | 6 | ABK87286 |
| 37 | 1446.8 | 30.1 | 12980 | 8 | ACA62469 |
| 38 | 1445.2 | 30.0 | 3423 | 2 | AAQ11083 |
| 39 | 1445.2 | 30.0 | 5201 | 12 | ADL72229 |
| 40 | 1444.8 | 30.0 | 2315 | 10 | ADF44791 |
| 41 | 1444.8 | 30.0 | 2600 | 10 | ADF44790 |
| 42 | 1444.8 | 30.0 | 4009 | 2 | AAQ70219 |
| 43 | 1442.8 | 30.0 | 3880 | 4 | AAF25570 |
| 44 | 1442.8 | 30.0 | 5115 | 2 | AAV81319 |
| 45 | 1442.8 | 30.0 | 5115 | 5 | AAH74745 |

ALIGNMENTS

RESULT 1
ACF58174
ID ACF58174 standard; DNA; 4814 BP.
AC ACF58174;
XX
DT 15-JAN-2004 (first entry)
DE Modified Tn7-Kan(Stop)-transposon.
XX
KW Transposon; genetic engineering; transposase; Cat-Mu; Tn7-Kan; ds.
XX OS Synthetic.
XX WO2003087370-A1.
XX 23-OCT-2003.
XX 14-APR-2003; 2003WO-FI000285.
XX 18-APR-2002; 2002FI-00000746.
XX (FINN-) FINNZYMES OV.
XX Savilahti H, Tieaho V;
XX WPI; 2003-845329/78.

New transposon nucleic acid comprising a genetically engineered translation stop signal within a transposon end sequence recognized by a transposase useful for producing deletion derivatives of polypeptide.

Claim 7; Page 31-32; Opp; English.

The invention relates to a transposon nucleic acid comprising a genetically engineered translation stop signal in the three reading frames at least partly within a transposon end sequence recognized by a transposase. The transposon is useful for producing deletion derivatives of polypeptide coding nucleic acids. The method involves performing a transposition reaction in the presence of a target nucleic acid containing a polypeptide coding nucleic acid and in the presence of a transposon containing a genetically engineered translation stop signal

| | | | | | | | | | | |
|----|--|--|---------------|-----------|--------------|--|--|--|--|--|
| CC | sequence in three reading frames at least partly within a transposon and | | | | | | | | | |
| CC | sequence recognized by a transposase; and recovering a target nucleic | | | | | | | | | |
| CC | acid having the transposon incorporated in the protein coding nucleic | | | | | | | | | |
| CC | acid. The present sequence represents a modified Tn7-Kan(Stop) - | | | | | | | | | |
| CC | transposon. | | | | | | | | | |
| XX | | | | | | | | | | |
| SQ | Sequence 4814 BP; 1214 A; 1206 C; 1137 G; 1257 T; 0 U; 0 Other; | | | | | | | | | |
| | Query Match | 100.0%; | Score 4814; | DB 10; | Length 4814; | | | | | |
| | Best Local Similarity | 100.0%; | Pred. No. 0; | | | | | | | |
| | Matches 4814; | Conservative 0; | Mismatches 0; | Indels 0; | Gaps 0; | | | | | |
| QY | 1 | GGTACCCTGTGAATGCGAAACCAACCCCTTGGCGAGACATATCATCGCGTCCGCCATCT | 60 | | | | | | | |
| DB | 1 | GGTACCCTGTGAATGCGCAAAACCAACCCCTTGGCGAGACATATCATCGCGTCCGCCATCT | 60 | | | | | | | |
| QY | 61 | CCAGCAGCGCACGGGGCGCATCTCGGGCAGCGTTGGTCTCGGCCACGGGTGCGCATGA | 120 | | | | | | | |
| DB | 61 | CCAGCAGCGCACGGGGCGCATCTCGGGCAGCGTTGGTCTCGGCCACGGGTGCGCATGA | 120 | | | | | | | |
| QY | 121 | TCGTGCTCTCTGCTGTGAGGACCCGGCTAGGCTCGCGGGGTTCCTTACTCGTTAGCAGA | 180 | | | | | | | |
| DB | 121 | TCGTGCTCTCTGCTGTGAGGACCCGGCTAGGCTCGCGGGGTTCCTTACTCGTTAGCAGA | 180 | | | | | | | |
| QY | 181 | ATGAATCACCGATACGGAGCGAAACGTGAAGCGACTGCTGCTGCAAAACGCTCGCGACCT | 240 | | | | | | | |
| DB | 181 | ATGAATCACCGATACGGAGCGAAACGTGAAGCGACTGCTGCTGCAAAACGCTCGCGACCT | 240 | | | | | | | |
| QY | 241 | GAGCAACAAATGAATGCTTCCTCGGTTTCGGTTCCTCGTGTTCCTGAAGCTCGAAACGGGAAGT | 300 | | | | | | | |
| DB | 241 | GAGCAACAAATGAATGCTTCCTCGGTTTCGGTTCCTCGTGTTCCTGAAGCTCGAAACGGGAAGT | 300 | | | | | | | |
| QY | 301 | CAGCGCCTGCACCATTAATGTTTCGGGATCTATGTCGGGTGCGGAGAAAGAGTAATGAAA | 360 | | | | | | | |
| DB | 301 | CAGCGCCTGCACCATTAATGTTTCGGGATCTATGTCGGGTGCGGAGAAAGAGTAATGAAA | 360 | | | | | | | |
| QY | 361 | TGGCAGATCCCTGGCTTGTGTCCACAACCGGTTAAACCTTTAAAGCTTTTAAAGCCTTAT | 420 | | | | | | | |
| DB | 361 | TGGCAGATCCCTGGCTTGTGTCCACAACCGGTTAAACCTTTAAAGCTTTTAAAGCCTTAT | 420 | | | | | | | |
| QY | 421 | ATATTCTTTTTTTTTCTTATAAACTTAAAAACCTTAGAGGCTATTATTAAGTTCGCTGATTTAT | 480 | | | | | | | |
| DB | 421 | ATATTCTTTTTTTTTCTTATAAACTTAAAAACCTTAGAGGCTATTATTAAGTTCGCTGATTTAT | 480 | | | | | | | |
| QY | 481 | ATTAATTTTATTGTTTCAAACATGAGAGCTTAGTACGTGAAACATGAGAGCTTTAGTACGTT | 540 | | | | | | | |
| DB | 481 | ATTAATTTTATTGTTTCAAACATGAGAGCTTAGTACGTGAAACATGAGAGCTTTAGTACGTT | 540 | | | | | | | |
| QY | 541 | AGCCATGAGAGCTTAGTACGTTAGCCATGAGGGTTTAGTTCGTTAAACATGAGAGCTTAG | 600 | | | | | | | |
| DB | 541 | AGCCATGAGAGCTTAGTACGTTAGCCATGAGGGTTTAGTTCGTTAAACATGAGAGCTTAG | 600 | | | | | | | |
| QY | 601 | TACGTTAAACATGAGAGCTTAGTACGTGAAACATGAGAGCTTAGTACGTACTATCAACAG | 660 | | | | | | | |
| DB | 601 | TACGTTAAACATGAGAGCTTAGTACGTGAAACATGAGAGCTTAGTACGTACTATCAACAG | 660 | | | | | | | |
| QY | 661 | GTTGAACTGCTGATCTTCGGAATCTATGTCGGGTGCGGAGAAAGAGTAATGAAATGGCAG | 720 | | | | | | | |
| DB | 661 | GTTGAACTGCTGATCTTCGGAATCTATGTCGGGTGCGGAGAAAGAGTAATGAAATGGCAG | 720 | | | | | | | |
| QY | 721 | ATCCCTGGCTTGTGTGCCAACCCGTTAAACCTTTAAAGCTTTTAAAGCCTTTATATATTC | 780 | | | | | | | |
| DB | 721 | ATCCCTGGCTTGTGTGCCAACCCGTTAAACCTTTAAAGCTTTTAAAGCCTTTATATATTC | 780 | | | | | | | |
| QY | 781 | TTTTTTTTTCTTATAAAACCTTTAAAGCTTTAGAGGCTATTATTAAGTTCGCTGATTTATTAAT | 840 | | | | | | | |
| DB | 781 | TTTTTTTTTCTTATAAAACCTTTAAAGCTTTAGAGGCTATTATTAAGTTCGCTGATTTATTAAT | 840 | | | | | | | |
| QY | 841 | TTTTTTTGTTCAAACATGAGAGCTTAGTACGTGAAACATGAGAGCTTAGTACGTAGCCAT | 900 | | | | | | | |
| DB | 841 | TTTTTTTGTTCAAACATGAGAGCTTAGTACGTGAAACATGAGAGCTTAGTACGTAGCCAT | 900 | | | | | | | |
| QY | 901 | GAGAGCTTAGTACGTTAGCCATGAGGGTTTAGTTCGTTTAAACATGAGAGCTTAGTACGTT | 960 | | | | | | | |

Db 1981 GTCCAGCGATGCCCTTGAGAGCCTTCAACCAGTCAGCTCCTTCGGTGGGCGCGGGCA 2040
Qy 2041 TGACTATCGTCCCGCACTTATGACTGTCTTTTATCATGCAACTCGTAGACAGAGTGC 2100
Db 2041 TGACTATCGTCCCGCACTTATGACTGTCTTTTATCATGCAACTCGTAGACAGAGTGC 2100
Qy 2101 CGGAGCGCTCTGGGTCACTTTTCGGCGAGGACCGCTTCGCTGGAGCGGAGCATGATG 2160
Db 2101 CGGAGCGCTCTGGGTCACTTTTCGGCGAGGACCGCTTCGCTGGAGCGGAGCATGATG 2160
Qy 2161 GCCTGTGCTTGGCGTATTCGGAATCTTGCAGCGCCTCGCTCAAGCCCTTCGCTCACTG 2220
Db 2161 GCCTGTGCTTGGCGTATTCGGAATCTTGCAGCGCCTCGCTCAAGCCCTTCGCTCACTG 2220
Qy 2221 CGGCCACAAAGCTTTTCGGCGAGAGAGAGGCCAATATCGCGGCATTCGGCGGAGCGGC 2280
Db 2221 CGGCCACAAAGCTTTTCGGCGAGAGAGAGGCCAATATCGCGGCATTCGGCGGAGCGGC 2280
Qy 2281 TGGGCTACGCTCTGCTGGGTTTCGGAGCGCGAGGCTGGATGGCCCTTCCCATATGATTC 2340
Db 2281 TGGGCTACGCTCTGCTGGGTTTCGGAGCGCGAGGCTGGATGGCCCTTCCCATATGATTC 2340
Qy 2341 TTCTCGCTTCGGCGGCATTCGGGATGCCCGCTTCAGGCCCATGCTGCCAGGCAAGTAG 2400
Db 2341 TTCTCGCTTCGGCGGCATTCGGGATGCCCGCTTCAGGCCCATGCTGCCAGGCAAGTAG 2400
Qy 2401 ATGACGACCATCAGGACAGCTTCAAGGATCGCTCGCGCTTACACAGCCTAACCTCGA 2460
Db 2401 ATGACGACCATCAGGACAGCTTCAAGGATCGCTCGCGCTTACACAGCCTAACCTCGA 2460
Qy 2461 TCATTGGACCGCTGATCGTCAGGCGATTTATCGCGCCTCGCGAGCACATGGAACGGGT 2520
Db 2461 TCATTGGACCGCTGATCGTCAGGCGATTTATCGCGCCTCGCGAGCACATGGAACGGGT 2520
Qy 2521 TGGCATGATTTAGCGCGCGCTTATACCTTGTCTGCTTCCCGCTTGGTCCGGTG 2580
Db 2521 TGGCATGATTTAGCGCGCGCTTATACCTTGTCTGCTTCCCGCTTGGTCCGGTG 2580
Qy 2581 CATGAGCGCGGCCACCTCGAATGGAAGCGCGGCGCACCTCGCTAACGGATTAC 2640
Db 2581 CATGAGCGCGGCCACCTCGAATGGAAGCGCGGCGCACCTCGCTAACGGATTAC 2640
Qy 2641 CACTCAAGATTTGAGCGCAATCAATTTCTGGGAGAACTGTAATGGGAAAACCAACCC 2700
Db 2641 CACTCAAGATTTGAGCGCAATCAATTTCTGGGAGAACTGTAATGGGAAAACCAACCC 2700
Qy 2701 TTGGCAGAACATATCCATCGGTCGCGCATCTCCAGAGCGCGACGCGCGCATCTCGGG 2760
Db 2701 TTGGCAGAACATATCCATCGGTCGCGCATCTCCAGAGCGCGACGCGCGCATCTCGGG 2760
Qy 2761 CAGCGTTGGGCTCGGGCTGGCATTTGACCTGAGTGATTTTCTCTGCTCCCGCGCATC 2820
Db 2761 CAGCGTTGGGCTCGGGCTGGCATTTGACCTGAGTGATTTTCTCTGCTCCCGCGCATC 2820
Qy 2821 CATACCGCAGTTGTTTACCTTCAAGCTTCCAGTAAACCGGCGATGTTCAATCATCAGTA 2880
Db 2821 CATACCGCAGTTGTTTACCTTCAAGCTTCCAGTAAACCGGCGATGTTCAATCATCAGTA 2880
Qy 2881 ACCGCTATCGTGAGCATCTCTCGTTTCAATCGGTATCATTTACCCCATGACAGAAAT 2940
Db 2881 ACCGCTATCGTGAGCATCTCTCGTTTCAATCGGTATCATTTACCCCATGACAGAAAT 2940
Qy 2941 CCCCCCTACAGGAGGCATCAGTACCAACAGGAAAAACCGCCCTTAAACATGGCCCGC 3000
Db 2941 CCCCCCTACAGGAGGCATCAGTACCAACAGGAAAAACCGCCCTTAAACATGGCCCGC 3000
Qy 3001 TTTATCAGAGCCAGACATTAACGCTTCTGGAGAACTCAACGAGCTGGACCGGATGAA 3060
Db 3001 TTTATCAGAGCCAGACATTAACGCTTCTGGAGAACTCAACGAGCTGGACCGGATGAA 3060
Qy 3061 CAGGAGAGCTTTACTCTCATGCCATCGGTATGAGTGAATAGATTAAGTCTTAAACTG 3120
Db 3061 CAGGAGAGCTTTACTCTCATGCCATCGGTATGAGTGAATAGATTAAGTCTTAAACTG 3120

Qy 3121 AACAAATAGATCTTAAACTATGCAAAATAAGTCTTTAAACTAGACAGAAATAGTTGTAAC 3180
Db 3121 AACAAATAGATCTTAAACTATGCAAAATAAGTCTTTAAACTAGACAGAAATAGTTGTAAC 3180
Qy 3181 GAAATCAGTCAGTCTGCTGTAAGAAAGCATACTGGACTTTTGTATGCGCTAAAGCAAA 3240
Db 3181 GAAATCAGTCAGTCTGCTGTAAGAAAGCATACTGGACTTTTGTATGCGCTAAAGCAAA 3240
Qy 3241 CTCTTCACTTTCTGAAGTGCAAAATTCGCCGCTGTAATTAAGAGGGCGCTGGGGTCGACGC 3300
Db 3241 CTCTTCACTTTCTGAAGTGCAAAATTCGCCGCTGTAATTAAGAGGGCGCTGGGGTCGACGC 3300
Qy 3301 GGGCGCTAACTATTAACCGTCTTAAGGTAGCGAGTTTAAACGATATCGGATTCGGCGCGCG 3360
Db 3301 GGGCGCTAACTATTAACCGTCTTAAGGTAGCGAGTTTAAACGATATCGGATTCGGCGCGCG 3360
Qy 3361 CTGAGGCTCGCTCGTGAAGAGGTTGCTGACTCATACGAGGCTCAATCGCCCAATC 3420
Db 3361 CTGAGGCTCGCTCGTGAAGAGGTTGCTGACTCATACGAGGCTCAATCGCCCAATC 3420
Qy 3421 ATCCAGCCAGAAAGTGAGGAGCCACGGTTGATGAGAGCTTTGTTGTAGGTGGAACAGTT 3480
Db 3421 ATCCAGCCAGAAAGTGAGGAGCCACGGTTGATGAGAGCTTTGTTGTAGGTGGAACAGTT 3480
Qy 3481 GGTGATTTTGAACCTTTTCTGCTTTGCCACGGAAACGGTCTGCGTTGTCCGGAAGATGGTGT 3540
Db 3481 GGTGATTTTGAACCTTTTCTGCTTTGCCACGGAAACGGTCTGCGTTGTCCGGAAGATGGTGT 3540
Qy 3541 CTGATCTCTCACTCAGCAAGAGTTGATTTATTCACAAAGCGCGCTCCGCTCAAGTC 3600
Db 3541 CTGATCTCTCACTCAGCAAGAGTTGATTTATTCACAAAGCGCGCTCCGCTCAAGTC 3600
Qy 3601 AGCGTAATGCTCTGCGAGTTTACAAACCAATTAACCAATTTCTGATTAGAAAACTCATCG 3660
Db 3601 AGCGTAATGCTCTGCGAGTTTACAAACCAATTAACCAATTTCTGATTAGAAAACTCATCG 3660
Qy 3661 AGCATCAAAATGAAACTGCAATTTATTCATATCAGGATTTATCAATACCATATTTTGA 3720
Db 3661 AGCATCAAAATGAAACTGCAATTTATTCATATCAGGATTTATCAATACCATATTTTGA 3720
Qy 3721 AGCGTTTCTGTAATGAGGAGAAACCTCAGGAGGAGTTCCATAGATGGCAGATCC 3780
Db 3721 AGCGTTTCTGTAATGAGGAGAAACCTCAGGAGGAGTTCCATAGATGGCAGATCC 3780
Qy 3781 TGGTATCGGTCTGCGATTCGACTCGTCCAACTCAATACCACTTAAATTTCCCTCG 3840
Db 3781 TGGTATCGGTCTGCGATTCGACTCGTCCAACTCAATACCACTTAAATTTCCCTCG 3840
Qy 3841 TCAAAAATTAAGGTTATCAAGTGAGAAATCACCATGAGTGACGATCGATTCGGTGAGAT 3900
Db 3841 TCAAAAATTAAGGTTATCAAGTGAGAAATCACCATGAGTGACGATCGATTCGGTGAGAT 3900
Qy 3901 GGGAAAAGCTTATGCAATTTCTTCCAGACTTTTCAACAGCGCCAGCATTAAGCTCGTCA 3960
Db 3901 GGGAAAAGCTTATGCAATTTCTTCCAGACTTTTCAACAGCGCCAGCATTAAGCTCGTCA 3960
Qy 3961 TCAAAATCAGTCGATCAACCAACCGTTATTCATTCGTGATTCGCTGAGCGAGCAGCA 4020
Db 3961 TCAAAATCAGTCGATCAACCAACCGTTATTCATTCGTGATTCGCTGAGCGAGCAGCA 4020
Qy 4021 AATACGCGATCGCTGTTTAAAGGACAAATTACAAACAGGAATCGAATGCAACCGCGCAGG 4080
Db 4021 AATACGCGATCGCTGTTTAAAGGACAAATTACAAACAGGAATCGAATGCAACCGCGCAGG 4080
Qy 4081 AACACTGCGAGCGATCAACAAATTTTACCTGAATCAGGATATTTCTTAATACCTGG 4140
Db 4081 AACACTGCGAGCGATCAACAAATTTTACCTGAATCAGGATATTTCTTAATACCTGG 4140
Qy 4141 AATCGTGTTCCTCGGGATCGAGTGTGAGTAACTGATCATCAGGAGTACGATA 4200
Db 4141 AATCGTGTTCCTCGGGATCGAGTGTGAGTAACTGATCATCAGGAGTACGATA 4200

Db 11786 TTGCTACTTGGAGCACATATCGACTACCGGATCATGGGCAACACCCGTCCTGTGGAT 11845
QY CCTTACCGCGGACGATCGTGGCCGGCATCAACCGGCCACAGGTGGGTTCTGTGGCG 1768
Db 11846 CCTTACGCGGACGATCGTGGCCGGCATCAACCGGCCACAGGTGGGTTCTGTGGCG 11905
QY CTATATCGCGACATCACCGATGGGAGAGATGGGCTGCCACTTCGGGCTCATGAGCG 1828
Db 11906 CTATATCGCGACATCACCGATGGGAGAGATGGGCTGCCACTTCGGGCTCATGAGCG 11965
QY TTGTTTGGCGTGGGTATCGTGGCGAGCCCGGTGGCCGGGAGACTGTGTGGCGCCATCTC 1888
Db 11966 TTGTTTGGCGTGGGTATCGTGGCGAGCCCGGTGGCCGGGAGACTGTGTGGCGCCATCTC 12025
QY TTGTCATGACCAATTCCTTGGCGCGCGGTGTCTCAAGGCGCTCAACCTACTACTGGGCTG 1948
Db 12026 TTGTCATGACCAATTCCTTGGCGCGCGGTGTCTCAAGGCGCTCAACCTACTACTGGGCTG 12085
QY TTTCCTTAATGACGAGATGCGATAAGGAGAGCGTGCACCGATGCCCTTGAGAGCCTTCA 2008
Db 12086 TTTCCTTAATGACGAGATGCGATAAGGAGAGCGTGCACCGATGCCCTTGAGAGCCTTCA 12145
QY CCAGTCAGCTCCTTCCGCTGGCGCGCGGCGATGACTATCGTCGCGGCACTTATGACTGT 2068
Db 12146 CCAGTCAGCTCCTTCCGCTGGCGCGCGGCGATGACTATCGTCGCGGCACTTATGACTGT 12205
QY TTCTTTTATCATGCAACTCTGTAGGACAGGTGCGGCGAGCGTCTGGGTCAATTTTCGGGCA 2128
Db 12206 TTCTTTTATCATGCAACTCTGTAGGACAGGTGCGGCGAGCGTCTGGGTCAATTTTCGGGCA 12265
QY GGACGCTTTTCTGCGAGCGGACAGATGATCGGCTGTGCTGCGGTATTCGGAATCTT 2188
Db 12266 GGACGCTTTTCTGCGAGCGGACAGATGATCGGCTGTGCTGCGGTATTCGGAATCTT 12325
QY GCAGCCCTCGCTCAAGCTTTCGTCACCTGCTCCGCGCACCAACCTTTCGGCGAGAACA 2248
Db 12326 GCAGCCCTCGCTCAAGCTTTCGTCACCTGCTCCGCGCACCAACCTTTCGGCGAGAACA 12385
QY GGCCATTATCGCGGCATGCGCGCGGACGCGCTGGGCTACGTCCTTCTGGCGGCTTCGCGAC 2308
Db 12386 GGCCATTATCGCGGCATGCGCGCGGACGCGCTGGGCTACGTCCTTCTGGCGGCTTCGCGAC 12445
QY GCGAGCTGGATGGCTTCCCATTAATGATTTCTGCTTCCGCGGCGATCGGATGCC 2368
Db 12446 GCGAGCTGGATGGCTTCCCATTAATGATTTCTGCTTCCGCGGCGATCGGATGCC 12505
QY CGCGTTGAGGCGCATGCTGTCCAGGACAGTACGACCCATCAGGACAGCTTCAAGG 2428
Db 12506 CGCGTTGAGGCGCATGCTGTCCAGGACAGTACGACCCATCAGGACAGCTTCAAGG 12565
QY ATCGCTCGCGGCTCTTACAGGCTAACTTCGATCATTTGACCGCTGATGTCACGCGCAT 2488
Db 12566 ATCGCTCGCGGCTCTTACAGGCTAACTTCGATCATTTGACCGCTGATGTCACGCGCAT 12625
QY TTATGCCGCTCGCGGACACATGAAACGGGTGGCATGATGATGAGCGCGGCGCTTATA 2548
Db 12626 TTATGCCGCTCGCGGACACATGAAACGGGTGGCATGATGATGAGCGCGGCGCTTATA 12685
QY CTTGTCTGCTTCCCGGCTTGGCTGCGGTGTCATGAGACCGGCGCACTTCGACCTGAAT 2608
Db 12686 CTTGTCTGCTTCCCGGCTTGGCTGCGGTGTCATGAGACCGGCGCACTTCGACCTGAAT 12745
QY GGAAGCCGGCGGACCTCGCTAAGGATTTCAACCTTCAAGAAATTTGGAGGCAATCAATTC 2668
Db 12746 GGAAGCCGGCGGACCTCGCTAAGGATTTCAACCTTCAAGAAATTTGGAGGCAATCAATTC 12805
QY TTGCGGAGAACTGTGAATGCGCAACCAACCCCTTGGCAGAACATATCCATCGGTCGCGC 2728
Db 12806 TTGCGGAGAACTGTGAATGCGCAACCAACCCCTTGGCAGAACATATCCATCGGTCGCGC 12865
QY ATCTTCAGACGCGCACCGGCGCATCTCGGCGAGCGTGGGTCTCTGG----- 2776

Db 12866 ATCTCCAGCAGCGGCACGCGGCGCATCTCGGCGAGGTTGGGTCTCTGGCCACGGGTGCGC 12925
QY 2777 ----- 2776
Db 12926 ATGATGTGCTCTCTGTGTGAGGACCGGCTAGGCTGGCGGGTTGCCTTACTGTTAG 12985
QY 2777 ----- 2776
Db 12986 CAGAATGAATCACCAGATAGCGAGCGAACGTAAGCGACTGCTGTGCAAAACGTCGTGG 13045
QY 2777 ----- 2776
Db 13046 ACCTGAGCAACAACATGAATGTCTTCGGTTCCGTTTCCGTTAAAGTCGTGNAACGCGG 13105
QY 2777 ----- 2776
Db 13106 AAGTCAGCGCCCTGCACCATATATGTTCGGATCTGCATCGCAGGATGCTGCTGGCTACCC 13165
QY 2777 -----CCTGGCATTTGACCTGAGTGAATTTTCTC 2805
Db 13166 TGTGAAACACCTPACATCTGTATTAAAGACGCGCTGGCATTTGACCTGAGTGAATTTTCTC 13225
QY TGGTCCCGCGCATCCATACCGCCAGTTGTTTACCTCAACAAGTTTCCAGTAACCGGGCA 2865
Db 13226 TGGTCCCGCGCATCCATACCGCCAGTTGTTTACCTCAACAAGTTTCCAGTAACCGGGCA 13285
QY TGTTCATCATCAAGTAACCGGTATCGTAGCATCTCTCTCGTTTTCATCGGTATCATTTACC 2925
Db 13286 TGTTCATCATCAAGTAACCGGTATCGTAGCATCTCTCTCGTTTTCATCGGTATCATTTACC 13345
QY CCATGAACAGAAATCCCTTTACAGGAGGATCAGTAGCAACAAACAGGAAAAACCGGC 2985
Db 13346 CCATGAACAGAAATCCCTTTACAGGAGGATCAGTAGCAACAAACAGGAAAAACCGGC 13405
QY CTTTACATGCGCCGCTTTTATCAGAGCCAGACATTAACGTTTCTCGAGAACTCAACGAG 3045
Db 13406 CTTTACATGCGCCGCTTTTATCAGAGCCAGACATTAACGTTTCTCGAGAACTCAACGAG 13465
QY 3046 CTGAGCGGATGAACAGGAGAGCTCT 3073
Db 13466 CTGAGCGGATGAACAGGAGAGCATCT 13493

RESULT 3

ADG93313
ID ADG93313 standard; DNA; 15159 BP.

XX AC ADG93313;

XX DT 11-MAR-2004 (first entry)

XX DE DEN2 (Tonga/74) cDNA plasmid P2.

XX immunogenic composition; dengue type 1; dengue type 2; dengue type 3;
dengue type 4; virucide; immunostimulant; vaccine; tetravalent vaccine;
dengue virus; delta30; attenuating mutation; humoral response;
cellular response; non-structural protein; structural protein;
dengue virus serotype; gene; ds; plasmid P2.

OS Dengue virus type 2.

XX WO2003092592-A2.

XX PD 13-NOV-2003.

XX PF 25-APR-2003; 2003WO-US013279.

XX PR 03-MAY-2002; 2002US-0377860P.

XX PR 23-DEC-2002; 2002US-0436500P.

XX (USSH) US DEPT HEALTH & HUMAN SERVICES.

XX Whitehead SS, Murphy BR, Markoff L, Falgout B, Blaney J;

Qy 2777 ----- 2776
 Db 12932 ATGATCGTGTCTCTGCTGTTGAGGACCCGGCTAGGCTGGCGGGTGTGCTTACTGGTTAG 12991
 Qy 2777 ----- 2776
 Db 12992 CAGATGAATCACCAGATAGCGGAGCGACGTGAAGCGACTGCTGCTGCAAAACGTCTGCG 13051
 Qy 2777 ----- 2776
 Db 13052 ACCTGAGCAACAACATGAATGCTCTTCGTTTCGTTGCTGTAAGTCTGGAACGCG 13111
 Qy 2777 ----- 2776
 Db 13112 AAGTCAGGCGCTGCACCATTAATGTTCCGGATCTGCATCGCAGGATGCTGCTGTACCC 13171
 Qy 2777 ----- 2805
 Db 13172 TGTGGAACACCTACATCTGTATTAAACGAAGCGCTGGCAATTGACCTGAGTGAATTTTCTC 13231
 Qy 2806 TGGTCCGCGCATCCATACCGCAGTGTGTTTACCTCACAACGTTCCAGTAAACCGGCA 2865
 Db 13232 TGGTCCGCGCATCCATACCGCAGTGTGTTTACCTCACAACGTTCCAGTAAACCGGCA 13291
 Qy 2866 TGTTCATCATCAGTAACCGGTATCGTGAGCATCTCTCTGTTTCATCGGTATCATTTACC 2925
 Db 13292 TGTTCATCATCAGTAACCGGTATCGTGAGCATCTCTCTGTTTCATCGGTATCATTTACC 13351
 Qy 2926 CCCATGAACAGAAATCCCTTTACACGAGGAGCATCAGTGACCAACAGGAAACACGCG 2985
 Db 13352 CCCATGAACAGAAATCCCTTTACACGAGGAGCATCAGTGACCAACAGGAAACACGCG 13411
 Qy 2986 CTTAAATGCGCCGCTTTATCAGAAAGCAGACATTAACGCTTCTGGAAGAACTCAACGAG 3045
 Db 13412 CTTAAATGCGCCGCTTTATCAGAAAGCAGACATTAACGCTTCTGGAAGAACTCAACGAG 13471
 Qy 3046 CTGACGCGGATGAACAGGAGAGCTCT 3073
 Db 13472 CTGACGCGGATGAACAGGAGAGCATCT 13499

RESULT 4

AEA06371
 ID AEA06371 standard; DNA; 7311 BP.
 XX AC AEA06371;
 XX 14-JUL-2005 (first entry)
 XX Novel bacterial host cell-related plasmid pCLO43 DNA SeqID7.
 DE gene expression; protein production; chromosome; genome; cell culture;
 XX antibiotic-resistance; ds.
 XX Unidentified.
 XX WO2005042750-A1.
 XX 12-MAY-2005.
 XX 29-OCT-2004; 2004WO-DK000750.
 XX 31-OCT-2003; 2003DK-00001624.
 XX (NOVO) NOVOZYMES AS.
 XX Olsen C, Rasmussen MD;
 XX WPI; 2005-346878/35.
 XX New bacterial host cell comprising at least two copies of an
 PT amplification unit in its genome, useful for increased production of a
 PT protein with several copies of a gene of interest stably chromosomally

PT integrated.

Example; SEQ ID NO 7; 53pp; English.

This invention relates to a novel bacterial host cell comprising at least two copies of an amplification unit in its genome, where two or more amplified chromosomal copies of the gene of interest are produced. The amplification unit comprises at least one copy of a gene of interest and an expressible conditionally essential gene, where the conditionally essential gene is either promoterless or transcribed from a heterologous promoter having an activity substantially lower than the endogenous promoter of the conditionally essential gene, and where the conditionally essential gene if not functional would render the cell auxotrophic for at least one specific substance or unable to utilize one or more specific sole carbon source. The bacterial host cell is useful for producing of a protein having several copies of a gene of interest stably chromosomally integrated, without leaving antibiotic resistance marker genes in the strains. The present sequence is that of a plasmid DNA sequence which was used in the exemplification of the invention.

Sequence 7311 BP; 1926 A; 1706 C; 1736 G; 1943 T; 0 U; 0 Other;

Query Match 32.4%; Score 1560.6; DB 14; Length 7311;

Best Local Similarity 98.8%; Pred. No. 5.7e-283;

Matches 1572; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

Qy 1109 TGTGGAACACCTACATCTGTATTAAACGAAGCATTTATTGAAGCATTTATCAGGGTTATTGT 1168
 Db 5721 TGTGGAATCTCACTCTCTCTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGT 5780
 Qy 1169 CTGATGAGCGGATACATATTGTAATGTAATTTAGAAAAATAAACAATAGGGTTCCGGCG 1228
 Db 5781 CTCATGAGCGGATACATATTGTAATGTAATTTAGAAAAATAAACAATAGGGTTCCGGCG 5840
 Qy 1229 ACATTTCCCGGAAAGGCCACCTGAGCTCTTAAGAAACCATTTATTATCATGACATTAAAC 1288
 Db 5841 ACATTTCCCGGAAAGGCCACCTGAGCTCTTAAGAAACCATTTATTATCATGACATTAAAC 5900
 Qy 1289 TATAAAATAGGGGTATCAGAGGCCCTTTTCGTTCTCAAGAAATTTCTCATGTTTGACAGCT 1348
 Db 5901 TATAAAATAGGGGTATCAGAGGCCCTTTTCGTTCTCAAGAAATTTCTCATGTTTGACAGCT 5960
 Qy 1349 TATCATCGATAAGCTTTAATGCGGTAGTTTATCAAGTTAAATTTGCTAAACGAGTCAGGC 1408
 Db 5961 TATCATCGATAAGCTTTAATGCGGTAGTTTATCAAGTTAAATTTGCTAAACGAGTCAGGC 6020
 Qy 1409 ACCGTGTATGAATCTAAACATGCGCTCATCTCGGACCGCTCCGACCCCTGGATGC 1468
 Db 6021 ACCGTGTATGAATCTAAACATGCGCTCATCTCGGACCGCTCCGACCCCTGGATGC 6080
 Qy 1469 TGTAGGATAGGGTTGGTTTATGCGGTACTGCGGGGCTCTTGGGGATATCGTCCATTC 1528
 Db 6081 TGTAGGATAGGGTTGGTTTATGCGGTACTGCGGGGCTCTTGGGGATATCGTCCATTC 6140
 Qy 1529 CGACAGATCGCCAGTCATATGCGGTGCTGTAGCGCTATATGCGTTGATGCAATTTCT 1588
 Db 6141 CGACAGATCGCCAGTCATATGCGGTGCTGTAGCGCTATATGCGTTGATGCAATTTCT 6200
 Qy 1589 ATGGCACCGGTTCTCGGACACCTGTCGACCGCTTTGGCGCGCCGCTCCCTGCTCGC 1648
 Db 6201 ATGGCACCGGTTCTCGGACACCTGTCGACCGCTTTGGCGCGCCGCTCCCTGCTCGC 6260
 Qy 1649 TTGCTACTTTGGAGCCACTATCGACTACGCGATCATGGCGACACACCCGCTCTGTGGAT 1708
 Db 6261 TTGCTACTTTGGAGCCACTATCGACTACGCGATCATGGCGACACACCCGCTCTGTGGAT 6320
 Qy 1709 CCTCTACGCGGACGATCGTGGCCGCGATCACCGGCGCCACAGGTTGCTGTGGCG 1768
 Db 6321 CCTCTACGCGGACGATCGTGGCCGCGATCACCGGCGCCACAGGTTGCTGTGGCG 6380
 Qy 1769 CTATATCGCGGACATCACCAGTGGGGAAGATCGGGCTCGGCATCTCGGGCTCATGAGCG 1828
 Db 6381 CTATATCGCGGACATCACCAGTGGGGAAGATCGGGCTCGGCATCTCGGGCTCATGAGCG 6440

| | | | | |
|----|------|---|------------------------------------|------|
| Qy | 1829 | TTGTTTCGGCGTGGGTATGTTGGCAGACCCCGCTGGCGGGGGA | CTGTTGGCGCCATCTC | 1888 |
| Db | 6441 | TTGTTTCGGCGTGGGTATGTTGGCAGACCCCGCTGGCGGGGGA | CTGTTGGCGCCATCTC | 6500 |
| Qy | 1889 | CTTGCAATGACCAATTCCTTCGGGGGGGGTGCTCAACGGCCTCAA | CCCTACTACTCTGGGCTG | 1948 |
| Db | 6501 | CTTGCAATGACCAATTCCTTCGGGGGGGGTGCTCAACGGCCTCAA | CCCTACTACTCTGGGCTG | 6560 |
| Qy | 1949 | CTTCCCTAAATGCAGAGTCGCATTAAGGGAGAGCGTCGACCGATG | CCCTCTTGAGAGCCTTCAA | 2008 |
| Db | 6561 | CTTCCCTAAATGCAGAGTCGCATTAAGGGAGAGCGTCGACCGATG | CCCTCTTGAGAGCCTTCAA | 6620 |
| Qy | 2009 | CCCACTCAGCTCCCTTCGGTGGCGGGGGCATGA | CTATCTGTCGGCCACTTTATGACTGT | 2068 |
| Db | 6621 | CCCACTCAGCTCCCTTCGGTGGCGGGGGCATGA | CTATCTGTCGGCCACTTTATGACTGT | 6680 |
| Qy | 2069 | CTTCTTTATCATGCAACTCGTAGGACAGGTGCGCGGAGCGCTCTG | GGTCAATTTTCGGCGA | 2128 |
| Db | 6681 | CTTCTTTATCATGCAACTCGTAGGACAGGTGCGCGGAGCGCTCTG | GGTCAATTTTCGGCGA | 6740 |
| Qy | 2129 | GGACCGCTTTTCGCTGGAGCCGACGATGATCGGCTGT | CGCTTCGCGTATTCGGAATCTT | 2188 |
| Db | 6741 | GGACCGCTTTTCGCTGGAGCCGACGATGATCGGCTGT | CGCTTCGCGTATTCGGAATCTT | 6800 |
| Qy | 2189 | GCAACGCCCTTCGCTCAAGCCTTCGTC | ACTGFTCCGCCACCAACGTTTCGGCGAGGAAGCA | 2248 |
| Db | 6801 | GCAACGCCCTTCGCTCAAGCCTTCGTC | ACTGFTCCGCCACCAACGTTTCGGCGAGGAAGCA | 6860 |
| Qy | 2249 | GGCCATTAATCGCGGCATGCGCGCGACGCGCTGGGGCTAGCTCT | TGCTGTCGSCGTTTCGCGAC | 2308 |
| Db | 6861 | GGCCATTAATCGCGGCATGCGCGCGACGCGCTGGGGCTAGCTCT | TGCTGTCGSCGTTTCGCGAC | 6920 |
| Qy | 2309 | GGCAGGCTGGATGGCCCTTCCCCATTAATGATTC | TTCTCGCTTCGGCGGGCATTCGGGATGCC | 2368 |
| Db | 6921 | GGCAGGCTGGATGGCCCTTCCCCATTAATGATTC | TTCTCGCTTCGGCGGGCATTCGGGATGCC | 6980 |
| Qy | 2369 | CGGCTTCAGGCCATGCTGCTCAGGCGAGGTAGATGACGACCATC | AGGGACAGCTTCAAGG | 2428 |
| Db | 6981 | CGGCTTCAGGCCATGCTGCTCAGGCGAGGTAGATGACGACCATC | AGGGACAGCTTCAAGG | 7040 |
| Qy | 2429 | ATCGCTCGCGCTCTTACCAACGCTTAACTTCGATCATTTGGAC | CGCTGATCGTCACGGCGAT | 2488 |
| Db | 7041 | ATCGCTCGCGCTCTTACCAACGCTTAACTTCGATCATTTGGAC | CGCTGATCGTCACGGCGAT | 7100 |
| Qy | 2489 | TTATGCCGCTTCGGCGAGCATGGAAACGGGTTGGCATGGAATTGT | AGGCGCCGCCCTTATA | 2548 |
| Db | 7101 | TTATGCCGCTTCGGCGAGCATGGAAACGGGTTGGCATGGAATTGT | AGGCGCCGCCCTTATA | 7160 |
| Qy | 2549 | CCTTGTCTGCCCTCCCGGGTTGGTTCGGGTGCAATGGAGCCGGG | CCACCTTCGACCTGAAT | 2608 |
| Db | 7161 | CCTTGTCTGCCCTCCCGGGTTGGTTCGGGTGCAATGGAGCCGGG | CCACCTTCGACCTGAAT | 7220 |
| Qy | 2609 | GGAAAGCCGGGCACTTCGCTTAACCGGATTCACCACTCCAAGAT | TTGAGGCAATCAATTC | 2668 |
| Db | 7221 | GGAAAGCCGGGCACTTCGCTTAACCGGATTCACCACTCCAAGAT | TTGAGGCAATCAATTC | 7280 |
| Qy | 2669 | TTGCGGAGAACTGTGAATGCGCAACCAACC | 2699 | |
| Db | 7281 | TTGCGGAGAACTGTGAATGCGCAACCAACC | 7311 | |

RESULT 5

AAV32978/C

ID AAV32978 standard; DNA; 5926 BP.

XX

AC AAV32978;

DT 17-NOV-1998 (first entry)

XXX

DE Tn7 donor plasmid sequence.

xxiii

transposable element; DNA sequencing; genetic analysis; insertional mutagenesis; ss.

REPORT

Synthetic.

References

WO9837205-A1

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

27-AUG-1998.

20-FEB-1998;

20-FEB-1997;

11/10/2011

(CRAI/) CRAI()

Question

Craig NL;

WDT: 1000-AC

WPI; 1998-46

Transmembrane

transposition (efficient and

efficient and
and altering

and altering

Disclosure:

12/20/2017

The present

Tn7 encodes

that allows

specificity

ATP-regulator

specific and

into a DNA s

analysis by

by insertion

```
Qy 3567 GATTATTCAACAAAGCCGCTCCGTCAGCTAGCTGCTCTCCAGTGTACAA 3626
Db 5131 GATTATTCAACAAAGCCGCTCCGTCAGCTAGCTGCTCTCCAGTGTACAA 5072
Qy 3627 CCAATTAAACCAATCTCGATTAGAAAACCTCATCGAGCATCAAAATGAAAAGTCAATTTATT 3686
Db 5071 CCAATTAAACCAATCTCGATTAGAAAACCTCATCGAGCATCAAAATGAAAAGTCAATTTATT 5012
Qy 3687 CATATCAGGATATCAATPACCATATTTTGGAAAAGCCGTTTCTGTGTAATGAAGAGAAAA 3746
Db 5011 CATATCAGGATATCAATPACCATATTTTGGAAAAGCCGTTTCTGTGTAATGAAGAGAAAA 4952
Qy 3747 CTCACCGAGGAGTTCCATAGATGCGCAAGATCCTGGTATCGGTCTCGGATTCGCACTCG 3806
Db 4951 CTCACCGAGGAGTTCCATAGATGCGCAAGATCCTGGTATCGGTCTCGGATTCGCACTCG 4892
Qy 3807 TCCAACTCAATCAACACCTATTAAATTTCCCTCTCGTCAAAAATAAGAGTTATCAAGTGAGAA 3866
Db 4891 TCCAACTCAATCAACACCTATTAAATTTCCCTCTCGTCAAAAATAAGAGTTATCAAGTGAGAA 4832
Qy 3867 ATCACCATGAGTACGACTGAATCCGGTGAGAAATGGCAAAAGCTTATGCAATTTCTTTCCA 3926
Db 4831 ATCACCATGAGTACGACTGAATCCGGTGAGAAATGGCAAAAGCTTATGCAATTTCTTTCCA 4772
Qy 3927 GACTTGTTCACAGCCAGCCATTACGCTCGTCATCAAAATCACTCGCATCAACCAACC 3986
Db 4771 GACTTGTTCACAGCCAGCCATTACGCTCGTCATCAAAATCACTCGCATCAACCAACC 4712
Qy 3987 GTTATTCAATCGTATCGCTGAGCGAGACGAAATACGCGATCGCTGTTTAAAGAGACA 4046
Db 4711 GTTATTCAATCGTATCGCTGAGCGAGACGAAATACGCGATCGCTGTTTAAAGAGACA 4652
Qy 4047 ATTACAAACAGGAATCGAATCCAAACCGCGCAGGAACATCCCGCGCATCAACCAATTT 4106
Db 4651 ATTACAAACAGGAATCGAATCCAAACCGCGCAGGAACATCCCGCGCATCAACCAATTT 4592
Qy 4107 TTGACCTGGAATCAGGATATTTCTTAATACCTGGAATGCTGTTTCCGGGGATCGCAGT 4166
Db 4591 TTGACCTGGAATCAGGATATTTCTTAATACCTGGAATGCTGTTTCCGGGGATCGCAGT 4532
Qy 4167 GGTGAGTAACCATGCATCATCAGGATACGATGAAATGCTTGTGTCGGAAGGCGAT 4236
Db 4531 GGTGAGTAACCATGCATCATCAGGATACGATGAAATGCTTGTGTCGGAAGGCGAT 4472
Qy 4227 AAATTCGCTCAGCCAGTTTAGTCTGACCATCTCATCTGTAAACATCATTTGGCAACGCTACC 4286
Db 4471 AAATTCGCTCAGCCAGTTTAGTCTGACCATCTCATCTGTAAACATCATTTGGCAACGCTACC 4412
Qy 4287 TTTGCCATGTTTCAGAAAACACTCTGGCGATCGGGCTTCCCATACATCATGATGATTTGT 4346
Db 4411 TTTGCCATGTTTCAGAAAACACTCTGGCGATCGGGCTTCCCATACATCATGATGATTTGT 4352
Qy 4347 CGCACCTGATGTCGCGCATTTATCGGAGCCATTTATACCCATATAAATCAATCAAGCATCAT 4406
Db 4351 CGCACCTGATGTCGCGCATTTATCGGAGCCATTTATACCCATATAAATCAATCAAGCATCAT 4292
Qy 4407 GTTGGAAATTTAATCGCGCCTCGAGCAAGAGCGTTTCCCGTTGAATATGGCTCATAAACCC 4466
Db 4291 GTTGGAAATTTAATCGCGCCTCGAGCAAGAGCGTTTCCCGTTGAATATGGCTCATAAACCC 4232
Qy 4467 CTTGTATTAATGTAAGCAGACAGTTTTTATTGTTTCATGATGATATTTTTTATC 4526
Db 4231 CTTGTATTAATGTAAGCAGACAGTTTTTATTGTTTCATGATGATATTTTTTATC 4172
Qy 4527 TTGTGCAATGTAATCAGAGATTTTGAGACACACGTCGCTTACTAGGATCCGATATCA 4586
Db 4171 TTGTGCAATGTAATCAGAGATTTTGAGACACACGTCGCTTACTAGGATCCGATATCA 4112
Qy 4587 TTTAAATCTAGGGATAACAGGGTAATACTAGTGTGACCAACAGATGAAGTGAATCTAG 4646
Db 4111 -----GGTCGACCAACAGATGAAGTGAATCTAG 4083
Qy 4647 TTCCAAACTATTTTGTCAATTTTAAATTTTCGATATTAGCTTACGACGCTACACCCAGTTCC 4706
```

```
Db 4082 TTCCAAACTATTTTGTCAATTTTAAATTTTCGATATTAGCTTACGACGCTACACCCAGTTCC 4023
Qy 4707 CATCTATTTTGTCACTCTTCCCTAAATAATCTTTAAAAAACCCTATTTCCACCCCTCCAG 4766
Db 4022 CATCTATTTTGTCACTCTTCCCTAAATAATCTTTAAAAAACCCTATTTCCACCCCTCCAG 3963
Qy 4767 TTCCCAACTATTTTCTACTCACTCAC 4792
Db 3962 TTCCCAACTATTTTGTCCGCCACAC 3937

RESULT 6
AAD45060/c
ID AAD45060 standard; DNA; 5926 BP.
XX AC AAD45060;
XX AC AAD45060;
XX 27-DEC-2002 (first entry)
XX pEM delta transposon donor plasmid DNA.
XX TnsC transposition regulatory protein; transposase; DNA sequencing;
XX genetic analysis; gene expression; ds.
XX Unidentified.
XX US6420524-B1.
XX PN 16-JUL-2002.
XX PD 20-FEB-1998; 98US-00027169.
XX PF 20-FEB-1997; 97US-0037955P.
XX PR (UYJO ) UNIV JOHNS HOPKINS SCHOOL MEDICINE.
XX PA Craig NL;
XX PI WPI; 2002-634795/68.
XX DR Mutant TnsC transposition regulatory protein for use in DNA sequencing
XX PT for genetic analysis, and alteration of gene expression by insertion of a
XX PT given genetic sequence, comprises an alanine to valine substitution.
XX PS Example 1; Col 71-76; 74pp; English.
XX CC The invention relates to a mutant TnsC transposition regulatory protein.
XX CC This mutant protein activates a transposase in the absence of TnsD or
XX CC TnsE, which directs intermolecular transposition of a transposable
XX CC element and discriminates between immune and nonimmune targets, in a
XX CC manner which is characterised by reduced target site selectivity. It may
XX CC be used in DNA sequencing methods, for genetic analysis by insertional
XX CC mutagenesis and alteration of gene expression by insertion of a given
XX CC genetic sequence. The present DNA sequence is transposon donor plasmid
XX CC pEM delta, a pBR plasmid containing a kanamycin mfn7 element. This
XX CC sequence is used in the exemplification of the invention
XX SQ Sequence 5926 BP; 1476 A; 1473 C; 1463 G; 1514 T; 0 U; 0 Other;

Query Match 30.7%; Score 1477.2; DB 6; Length 5926;
Best Local Similarity 93.4%; Pred. No. 2.4e-267;
Matches 1594; Conservative 0; Mismatches 38; Indels 74; Gaps 2;

Qy 3087 TCCGTATGAGTGTAGTAATTAAGTCTTAACTCAACAAATAGATCTTAAACTATGACAA 3146
Db 5568 TCTAAATGTGGCGGCAATAAAGTCTTAACTCAACAAATAGATCTTAAACTATGACAA 5509
Qy 3147 TAAAGTCTTAAACTATGACAGAAATAGTGTAACTGAAATCAGTCCAGTTATGCTGTGAAA 3206
Db 5508 TAAAGTCTTAAACTATGACAGAAATAGTGTAACTGAAATCAGTCCAGTTATGCTGTGAAA 5449
Qy 3207 AAGCATACTGGACTTTTGTATTATGGCTAAAGCAAACTCTTCAATTTTCTGAAGTGCAAAATTG 3266
```

Db 5448 |||||AAGCACTACTGGACTTTTGTATGGCTAAAGCAAACTCTTCATTTCTGAAGTGCATATTG 5389
 Qy 3267 CCGTCTGATTAAGAGAGGGCGTGGGTGCGAGCGCGCGCTAACTATAACGGTCTTAAGG 3326
 Db 5388 CCGTCTGATTAAGAGAGGGCGTGGGTGCGAGCGCGCGCTAACTATAACGGTCTTAAGG 5357
 Qy 3327 TAGCGAGTTTAAACGATACGGATCCCGCCCGCCGCTGAGGTCTGCTCGTGAAGAGGTG 3386
 Db 5356 -----CTGCAAGGGGGGGGGGGCGCTGAGGTCTGCCCTCGTGAAGAGGTG 5312
 Qy 3387 TTGCTGACTCATACAGCGCTGAATCGCCCATCATCCAGCAGAAAAGTGAAGGAGCCAC 3446
 Db 5311 TTGCTGACTCATACAGCGCTGAATCGCCCATCATCCAGCAGAAAAGTGAAGGAGCCAC 5252
 Qy 3447 GGTGATGAGAGCTTTGTGTTGAGTGGAGCCAGTTGGTGAATTTGAACTTTTGTCTTGCCA 3506
 Db 5251 GGTGATGAGAGCTTTGTGTTGAGTGGAGCCAGTTGGTGAATTTGAACTTTTGTCTTGCCA 5192
 Qy 3507 CGGAACGGTCTGCGTTGTCGGGAAGATCGGTGATCTGATCCTTCAACTCAGCAAGAGTTC 3566
 Db 5191 CGGAACGGTCTGCGTTGTCGGGAAGATCGGTGATCTGATCCTTCAACTCAGCAAAAGTTC 5132
 Qy 3567 GATTTATTCAAAAGCGCGCTCCCGTCAAGTCAGGTAATGCTCTGCCAGTGTACAA 3626
 Db 5131 GATTTATTCAAAAGCGCGCTCCCGTCAAGTCAGGTAATGCTCTGCCAGTGTACAA 5072
 Qy 3627 CCAATTAAACCAATCTGATTAAGAAAACCTCATCAGAGCATCAAAATGAACTGCAATTTATT 3686
 Db 5071 CCAATTAAACCAATCTGATTAAGAAAACCTCATCAGAGCATCAAAATGAACTGCAATTTATT 5012
 Qy 3687 CATATCAGGATTAATCAATACCATATTTTGAAGAGCGTTTCTGTAATGAGGAGAANA 3746
 Db 5011 CATATCAGGATTAATCAATACCATATTTTGAAGAGCGTTTCTGTAATGAGGAGAANA 4952
 Qy 3747 CTCACGAGGCGAGTTCCATAGAGTGGCAAGATCCTGGTATCGGTCTGCGATTCCGACTCG 3806
 Db 4951 CTCACGAGGCGAGTTCCATAGAGTGGCAAGATCCTGGTATCGGTCTGCGATTCCGACTCG 4892
 Qy 3807 TCCAACTCAATACAACTATTAAATTTCCCTCGTCAAAAATAAGGTTATCAAGTGAGAA 3866
 Db 4891 TCCAACTCAATACAACTATTAAATTTCCCTCGTCAAAAATAAGGTTATCAAGTGAGAA 4832
 Qy 3867 ATCACCATGAGTACGACTGAATCCGTTGAGATGGCAAAAGCTTATGCAATTTCTTCCA 3926
 Db 4831 ATCACCATGAGTACGACTGAATCCGTTGAGATGGCAAAAGCTTATGCAATTTCTTCCA 4772
 Qy 3927 GACTTGTTCACAGGCCAGCCATTACGCTCGTCAAAAATCACTCGCATCAACCAAAACC 3986
 Db 4771 GACTTGTTCACAGGCCAGCCATTACGCTCGTCAAAAATCACTCGCATCAACCAAAACC 4712
 Qy 3987 GTTATTCATTCGTGATTCGCTGAGCGAGACGAAATACGCGATCGCTGTTTAAAGGACA 4046
 Db 4711 GTTATTCATTCGTGATTCGCTGAGCGAGACGAAATACGCGATCGCTGTTTAAAGGACA 4652
 Qy 4047 ATTACAAACAGGAATCGAATCGAACCGCGGAGGAACTGCGCAGCGCATCAACATATT 4106
 Db 4651 ATTACAAACAGGAATCGAATCGAACCGCGGAGGAACTGCGCAGCGCATCAACATATT 4592
 Qy 4107 TTCACCTGAATCAGGATATTCTTCTAATACCTGGAATGCTGTTTCCCGGGATCGCAAT 4166
 Db 4591 TTCACCTGAATCAGGATATTCTTCTAATACCTGGAATGCTGTTTCCCGGGATCGCAAT 4532
 Qy 4167 GGTGAGTAACCATGATCATCAGGAGTACGGAATAAAATCTTGTGTCGGAAGAGGCAT 4226
 Db 4531 GGTGAGTAACCATGATCATCAGGAGTACGGAATAAAATCTTGTGTCGGAAGAGGCAT 4472
 Qy 4227 AAATTCGTCAGCCAGTTTGTCTGACCATCTCATCTGTAACATCATTTGGCNAACGCTACC 4286
 Db 4471 AAATTCGTCAGCCAGTTTGTCTGACCATCTCATCTGTAACATCATTTGGCNAACGCTACC 4412
 Qy 4287 TTGTCGATGTTTTCAGAAAACACTCGGCGCATCGGGCTTCCCATCAATTCGATAGATTGT 4346
 |||||

Db 4411 TTTGCCATGTTTCAGAAAACAACTCTGGCGCATCGGGCTTCCCATACAATCGATAGATTGT 4352
 Qy 4347 CGCACCTGATTGCCCGACATTATCGCGAGCCCAATTTATATCCCATATAAATCAGCATCCAT 4406
 Db 4351 CGCACCTGATTGCCCGACATTATCGCGAGCCCAATTTATATCCCATATAAATCAGCATCCAT 4292
 Qy 4407 GTTGGAAATTTAATCGCGCCCTCGAGCAAGACGTTTCCCGTTGTAATATGGCTCATAACACC 4466
 Db 4291 GTTGGAAATTTAATCGCGCCCTCGAGCAAGACGTTTCCCGTTGTAATATGGCTCATAACACC 4232
 Qy 4467 CTTGTATTAATCTGTTATGTAAGCAGACAGTATTTTATGTTTCATGATGATATATTTTATC 4526
 Db 4231 CTTGTATTAATCTGTTATGTAAGCAGACAGTATTTTATGTTTCATGATGATATATTTTATC 4172
 Qy 4527 TTGTGCAATCTAATCATCAGAGATTTTGAGACACAACTGGCTTACTAGGATCCGATATCA 4586
 Db 4171 TTGTGCAATCTAATCATCAGAGATTTTGAGACACAACTGGCTTACTAGGATCCGATATCA 4112
 Qy 4587 TTTAAATCTAGGATTAACAGGTAATACTAGTGTGCAACCAACAGATAAGTGAATCTAG 4646
 Db 4111 -----GGTCGACCAACAGATAAGTGAATCTAG 4083
 Qy 4647 TTCCAAACTATTTTGTCAATTTTAAATTTTTCGTATTTAGCTTACGACGCTACACCCAGTTCC 4706
 Db 4082 TTCCAAACTATTTTGTCAATTTTAAATTTTTCGTATTTAGCTTACGACGCTACACCCAGTTCC 4023
 Qy 4707 CATCTATTTTGTCACTCTTCCCTAAATAATCCTTAAAACTCCATTTCCACCCCTCCAG 4766
 Db 4022 CATCTATTTTGTCACTCTTCCCTAAATAATCCTTAAAACTCCATTTCCACCCCTCCAG 3963
 Qy 4767 TTCCCAACTATTTTCTACTCACTCAC 4792
 Db 3962 TTCCCAACTATTTTGTTCGCCCCACAC 3937
 RESULT 7
 AAV32977/c
 ID AAV32977 standard; DNA; 5926 BP.
 AC AAV32977;
 XX
 DT 17-NOV-1998 (first entry)
 XX
 XX Tn7 donor plasmid sequence.
 DE Tn7; donor plasmid; transposon; ATP-utilising regulatory protein;
 XX transposable element; DNA sequencing; genetic analysis;
 KW insertional mutagenesis; ss.
 XX Synthetic.
 OS
 XX WO9837205-A1.
 FN
 XX
 PD 27-AUG-1998.
 XX
 PF 20-FEB-1998; 98WO-US003353.
 XX
 PR 20-FEB-1997; 97US-00037955.
 XX
 XX (CRAI/) CRAIG N L.
 PA
 XX Craig NL;
 PI
 XX WPI; 1998-467567/40.
 DR
 XX Transposon(s) encoding mutant ATP using proteins for insertion - which is
 PT efficient and random, with reduced site specificity; for DNA sequencing
 PT and altering gene expression.
 XX
 PS Disclosure; Fig 10B; 143pp; English.
 CC The present sequence represents a Tn7 donor plasmid pEW delta R.adj.1.
 CC The plasmid carries a 1625 bp mini-Tn7 element; 199bp of Tn7L flank a

kanamycin gene with SalI sites at the junctions. The backbone is pTRC99 (Pharmacia). Transposon Tn7 encodes an ATP-utilising regulatory protein that contains a mutation that allows efficient and simple insertion of, and reduced target site specificity of, a transposable element derived from the transposon. The ATP-regulatory protein and compositions are used for the efficient, non-specific and simple insertion of a transposon or transposable element into a DNA segment. This is useful in DNA sequencing, for genetic analysis by insertional mutagenesis, or for alteration of gene expression by insertion of a desired sequence

Sequence 5926 BP; 1476 A; 1473 C; 1463 G; 1514 T; 0 U; 0 Other;

Query Match 30.7%; Score 1476; DB 2; Length 5926;
Best Local Similarity 93.6%; Pred. No. 4.1e-267;
Matches 1591; Conservative 0; Mismatches 35; Indels 74; Gaps 2;

| | | | |
|----|------|--|------|
| Qy | 3093 | TGAGTGACGTAGATAAAGCTCTTAAACTGAAACAAAATAGATCTAAACTATGACAAATAAGT | 3152 |
| Db | 5926 | TGTGGCGGACATTAAGCTCTTAACTGAAACAAAATAGATCTAAACTATGACAAATAAGT | 5867 |
| Qy | 3153 | CTTAAACTAGACAGAAATAGTTGTAACCTGAAATCAGTCCAGTTATGCTGTGAAAGAGCAT | 3212 |
| Db | 5866 | CTTAAACTAGACAGAAATAGTTGTAACCTGAAATCAGTCCAGTTATGCTGTGAAAGAGCAT | 5807 |
| Qy | 3213 | ACTGACCTTTTGTATGCTAAGCAAACTCTTCAATTTCTGAAGTGAATTTGCCCGTC | 3272 |
| Db | 5806 | ACTGACCTTTTGTATGCTAAGCAAACTCTTCAATTTCTGAAGTGAATTTGCCCGTC | 5747 |
| Qy | 3273 | GTATTAAGAGGGGGCTGGGTCGACGGCGGCTTAATCAATCAATCGGTCTTAAGTAGGCA | 3332 |
| Db | 5746 | GTATTAAGAGGGGGCTGGGTCGAC----- | 5721 |
| Qy | 3333 | GTTTAAACGATATCGATCCGGCCCGCTGAGTCTCGTCTGGAAGAGTGTGCTG | 3392 |
| Db | 5720 | -----CTGCAGGGGGGGGGGGCGCTGAGTCTCGTCTGGAAGAGTGTGCTG | 5670 |
| Qy | 3393 | ACTCATACAGGCTGAATCGCCCATCATCCAGCCAGAAAGTGAAGGAGCCACGGTTGA | 3452 |
| Db | 5669 | ACTCATACAGGCTGAATCGCCCATCATCCAGCCAGAAAGTGAAGGAGCCACGGTTGA | 5610 |
| Qy | 3453 | TGAGAGCTTTGTTGAGTGACACAGTTGGTGAATTTTGAATTTTGGCTTCCACGGAGC | 3512 |
| Db | 5609 | TGAGAGCTTTGTTGAGTGACACAGTTGGTGAATTTTGAATTTTGGCTTCCACGGAGC | 5550 |
| Qy | 3513 | GGTCTGCGTTGTCGGGAAGATCGCTGATCTCATCTTCAATCTCAGCAAGAGTTCGATTTA | 3572 |
| Db | 5549 | GGTCTGCGTTGTCGGGAAGATCGCTGATCTCATCTTCAATCTCAGCAAGAGTTCGATTTA | 5490 |
| Qy | 3573 | TTCAACAAAGCGCGCTCCGTCAGGTAAATGCTCTGCCAGTGTTCACACCAAT | 3632 |
| Db | 5489 | TTCAACAAAGCGCGCTCCGTCAGGTAAATGCTCTGCCAGTGTTCACACCAAT | 5430 |
| Qy | 3633 | AACCAATTTCTAGTAAAGAACTCATCCAGATCAATGAACTGCAATTTATTCATATC | 3692 |
| Db | 5429 | AACCAATTTCTAGTAAAGAACTCATCCAGATCAATGAACTGCAATTTATTCATATC | 5370 |
| Qy | 3693 | AGGATTAATCAATACCATATTTTGAAGAAAGCGTTTCTGTAATGAAGGAGAAACTCAC | 3752 |
| Db | 5369 | AGGATTAATCAATACCATATTTTGAAGAAAGCGTTTCTGTAATGAAGGAGAAACTCAC | 5310 |
| Qy | 3753 | GAGGAGTTCCATAGGATGGAAGATCTGGTATCGGTCTCGGATTCGGACTCGTCCAAC | 3812 |
| Db | 5309 | GAGGAGTTCCATAGGATGGAAGATCTGGTATCGGTCTCGGATTCGGACTCGTCCAAC | 5250 |
| Qy | 3813 | ATCAATACAACTTATTAATTTCCCTCTGTCAAAATTAAGGTTATCAAGTGAAGAAATCAC | 3872 |
| Db | 5249 | ATCAATACAACTTATTAATTTCCCTCTGTCAAAATTAAGGTTATCAAGTGAAGAAATCAC | 5190 |
| Qy | 3873 | ATGAGTGACGACTGAATCCGGTGGAATGGCAAGCTTATGCAATTTCTTCCAGACTTG | 3932 |
| Db | 5189 | ATGAGTGACGACTGAATCCGGTGGAATGGCAAGCTTATGCAATTTCTTCCAGACTTG | 5130 |
| Qy | 3933 | TTCAACAGGGCGGCAATTAACGCTCGTCAATCAAAATCACTCGCATCAACCAACCGTTATT | 3992 |

| | | | |
|----|------|--|------|
| Db | 5129 | TTCAACAGGGCGGCAATTAACGCTCGTCAATCAAAATCACTCGCATCAACCAACCGTTATT | 5070 |
| Qy | 3993 | CATTCTGTCATTGCGCTCGAGCGAGCAAAATACGGGATCGCTGTTTAAAGAGCAATTACA | 4052 |
| Db | 5069 | CATTCTGTCATTGCGCTCGAGCGAGCAAAATACGGGATCGCTGTTTAAAGAGCAATTACA | 5010 |
| Qy | 4053 | AACAGGAATCGAATGCAACCGCGGAGCAACACATGCGGCGCATCAACAATATTTTCACC | 4112 |
| Db | 5009 | AACAGGAATCGAATGCAACCGCGGAGCAACACATGCGGCGCATCAACAATATTTTCACC | 4950 |
| Qy | 4113 | TGAATCAGGATATTTCTTAATACCTGGAATGCTTTTCCCGGGATCGCAGTGGTGAG | 4172 |
| Db | 4949 | TGAATCAGGATATTTCTTAATACCTGGAATGCTTTTCCCGGGATCGCAGTGGTGAG | 4890 |
| Qy | 4173 | TAAACCATCATCATCAGGAGTACGGATAAAATGCTTGAATGTCGGAAGGCAATAAATTC | 4232 |
| Db | 4889 | TAAACCATCATCATCAGGAGTACGGATAAAATGCTTGAATGTCGGAAGGCAATAAATTC | 4830 |
| Qy | 4233 | CGTCAGCCAGTTTATGTCGACCATCTCATCTGTAAATCATCATTTGGCAACGCTTGGC | 4292 |
| Db | 4829 | CGTCAGCCAGTTTATGTCGACCATCTCATCTGTAAATCATCATTTGGCAACGCTTGGC | 4770 |
| Qy | 4293 | ATGTTTTCAGAAACAACTCTGCGCATCGGGCTTCCCATCAATCGATAGATTGTCGCACC | 4352 |
| Db | 4769 | ATGTTTTCAGAAACAACTCTGCGCATCGGGCTTCCCATCAATCGATAGATTGTCGCACC | 4710 |
| Qy | 4353 | TGATTTGCCGACATTTATCGCGAGCCCAATTTATACCATATTAATCAGCATCCATCTTGA | 4412 |
| Db | 4709 | TGATTTGCCGACATTTATCGCGAGCCCAATTTATACCATATTAATCAGCATCCATCTTGA | 4650 |
| Qy | 4413 | ATTTAAATCGCGGCTCGAGCAAGACGTTTCCCGTTGAAATATGGCTCATAACACCCCTTGT | 4472 |
| Db | 4649 | ATTTAAATCGCGGCTCGAGCAAGACGTTTCCCGTTGAAATATGGCTCATAACACCCCTTGT | 4590 |
| Qy | 4473 | ATTTACTGTTTATGTAAGCAGACAGTTTATTTATTTGTTATGATATATATTTTATCTTTGTC | 4532 |
| Db | 4589 | ATTTACTGTTTATGTAAGCAGACAGTTTATTTATTTGTTATGATATATATTTTATCTTTGTC | 4530 |
| Qy | 4533 | AATGTAACATCAGAGATTTTTCAGACACAACTGCGTTACTAGGATCCGATATCATTTAAA | 4592 |
| Db | 4529 | AATGTAACATCAGAGATTTTTCAGACACAACTGCGTTACTAGGATCCGATATCATTTAAA | 4476 |
| Qy | 4593 | TCTAGGATTAACAGGGTAATACTAGTGTGCGACCAACAGATAAGTGAATCTAGTTCCAA | 4652 |
| Db | 4475 | -----GGTCCGACCAACAGATAAGTGAATCTAGTTCCAA | 4441 |
| Qy | 4653 | ACTATTTTGTCAATTTTAAATTTTTCGATATAGCTTACGACGCTACACCGATTTCCATCTA | 4712 |
| Db | 4440 | ACTATTTTGTCAATTTTAAATTTTTCGATATAGCTTACGACGCTACACCGATTTCCATCTA | 4381 |
| Qy | 4713 | TTTTGTCACTCTTCCCTAAATAATCCTTAAATACTCCATTTCCACCCCTCCAGTTCCCA | 4772 |
| Db | 4380 | TTTTGTCACTCTTCCCTAAATAATCCTTAAATACTCCATTTCCACCCCTCCAGTTCCCA | 4321 |
| Qy | 4773 | ACTATTTTCTACTCACTCAC 4792 | |
| Db | 4320 | ACTATTTTGTCCGCCACAC 4301 | |

RESULT 8

AAD45059/c

ID AAD45059 standard; DNA; 5926 BP.

XX AC AAD45059;

XX DT 27-DEC-2002 (first entry)

XX Transposon donor plasmid pEM delta R.adj to 1.

XX TnsC transposition regulatory protein; transposase; DNA sequencing;
XX genetic analysis; gene expression; ds.

Db 4440 ACTATTTTGTGCTATTTTAAATTTTGGTATTTAGCTTAGCGCTACACCGAGTTCCCATCTA 4381
 QY 4713 TTTTGTCACTCTCCCTAAATAATCTTAAATACTCAATTTCCACCCCTCCAGTTCCCA 4772
 Db 4380 TTTTGTCACTCTCTCCCTAAATAATCTTAAATACTCAATTTCCACCCCTCCAGTTCCCA 4321
 QY 4773 ACTATTTTCTACTCACTCAC 4792
 Db 4320 ACTATTTTGTGCGCCACAC 4301

RESULT 9
 ADG46817/c
 ID ADG46817 standard; DNA; 5926 BP.
 AC ADG46817;
 XX
 DT 11-MAR-2004 (first entry)
 DE Donor plasmid pEM delta R.adj to 1 DNA.

XX Transposon; ATP-utilising regulatory protein; genetic analysis;
 KW insertional mutagenesis; gene; plasmid ; cyclic; circular; ds.
 XX
 OS Unidentified.

XX US2002188105-A1.
 PN 12-DEC-2002.
 XX 19-DEC-2001; 2001US-00024809.
 XX 20-FEB-1997; 97US-0037955P.
 PR 20-FEB-1998; 98US-00027169.
 XX (UWJO) UNIV JOHNS HOPKINS SCHOOL MEDICINE.

XX Craig NL;
 PI WPI; 2004-020185/02.
 DR
 XX Transposon for use in e.g. DNA sequencing methods, comprises a mutation
 PT that allows efficient and simple insertion of and reduced target site
 PT specificity on the transposon.
 XX

PS Example 4; SEQ ID NO 3; 80pp; English.
 XX
 XX The invention relates to a transposon encoding an ATP-utilising
 CC regulatory protein containing a mutation that allows efficient and simple
 CC insertion of and reduced target site specificity on the transposon. The
 CC invention is used in DNA sequencing methods, for genetic analysis by
 CC insertional mutagenesis, and alteration of gene expression. The invention
 CC is used to achieve efficient, simple, non-specific or random insertion
 CC into any given DNA segment. The present sequence is donor plasmid pEM
 CC delta R.adj to 1 DNA.

XX Sequence 5926 BP; 1476 A; 1473 C; 1463 G; 1514 T; 0 U; 0 Other;
 Query Match 30.7%; Score 1476; DB 12; Length 5926;
 Best Local Similarity 93.6%; Pred. No. 4.1e-267;
 Matches 1591; Conservative 0; Mismatches 35; Indels 74; Gaps 2;

QY 3093 TGAGTGAGTAGAATAAAGTCTTAAACTGAAACAAATAGATCTAAACTATGACATAAAGT 3152
 Db 5926 TGTGGCGGACATAAAGTCTTAACTGAAACAAATAGATCTAAACTATGACATAAAGT 5867
 QY 3153 CTTAAACTAGACAGAAATAGTGTAAACTGAAATCAGTCCAGTTATGCTGTGAAATAAGCAT 3212
 Db 5866 CTTAAACTAGACAGAAATAGTGTAAACTGAAATCAGTCCAGTTATGCTGTGAAATAAGCAT 5807
 QY 3213 ACTGGACTTTTGTATGGCTAAGCAAACTCTTCAATTTCTGAAAGTCAAAATGCGCGTC 3272
 Db 5806 ACTGGACTTTTGTATGGCTAAGCAAACTCTTCAATTTCTGAAAGTCAAAATGCGCGTC 5747

QY 3273 GTATTAAAGAGGGGGCTGGGGTTCGACGGCGCGCTAACTATAAACGGTCTCTAAGGTAGCGA 3332
 Db 5746 GTATTAAAGAGGGGGCTGGGGTTCGAC----- 5721
 QY 3333 GTTTAAACGATATCGGATCCGGCCGCTGAGGTTCTGCTGTGAAGAAGGTGTTGCTG 3392
 Db 5720 -----CTGACGGGGGGGGGGGGCGCTGAGGTCTGCTGTGAAGAAGGTGTTGCTG 5670
 QY 3393 ACTCATACGAGGCTGAATTCGCCCATCATCAGCGAGAAAGTGAGGGAGCCACGTTGA 3452
 Db 5669 ACTCATACGAGGCTGAATTCGCCCATCATCAGCGAGAAAGTGAGGGAGCCACGTTGA 5610
 QY 3453 TGAGAGCTTTGTGTAGGTGGAACCAAGTTGGTGAATTTTGAATTTTGTGCTTTTGCACGGAAC 3512
 Db 5609 TGAGAGCTTTGTGTAGGTGGAACCAAGTTGGTGAATTTTGAATTTTGTGCTTTTGCACGGAAC 5550
 QY 3513 GGTCTGCTGTTGTCGGGAAGATCGGTGATCTTCACTCAGCAAGAGTTCGATTTA 3572
 Db 5549 GGTCTGCTGTTGTCGGGAAGATCGGTGATCTTCACTCAGCAAGAGTTCGATTTA 5490
 QY 3573 TTCAACAAAGCGCGCTCCCGTCAAGTCAGCGTAAATGCTCTGCCAGTGTTCACACCAATT 3632
 Db 5489 TTCAACAAAGCGCGCTCCCGTCAAGTCAGCGTAAATGCTCTGCCAGTGTTCACACCAATT 5430
 QY 3633 AACCAATTTCTGATTAGAAAACTCATCGAGCATCAAAATGAAACTGCAATTTTATTCATATC 3692
 Db 5429 AACCAATTTCTGATTAGAAAACTCATCGAGCATCAAAATGAAACTGCAATTTTATTCATATC 5370
 QY 3693 AGGATTATCAATACCATATTTTGAAGAGCGGTTTCTGTAATGAAGGAGAAATTCACC 3752
 Db 5369 AGGATTATCAATACCATATTTTGAAGAGCGGTTTCTGTAATGAAGGAGAAATTCACC 5310
 QY 3753 GAGCAGTTCCATAGGATGGAAGATCTGCTGATCGCTGCGATTCGACCTCGTCCAC 3812
 Db 5309 GAGCAGTTCCATAGGATGGAAGATCTGCTGATCGCTGCGATTCGACCTCGTCCAC 5250
 QY 3813 ATCAATACAACTATTAATTTTCCCTCGTCAAAATAAGGTTATCAAGTGAGAAATTCACC 3872
 Db 5249 ATCAATACAACTATTAATTTTCCCTCGTCAAAATAAGGTTATCAAGTGAGAAATTCACC 5190
 QY 3873 ATGAGTGAGCATGAAATCCGGTGAGATGGAAGAGCTTATGCAATTTCTTCCAGACTTG 3932
 Db 5189 ATGAGTGAGCATGAAATCCGGTGAGATGGAAGAGCTTATGCAATTTCTTCCAGACTTG 5130
 QY 3933 TTCAACAGCGCAGCATTAACGCTGTCATCAAAATCACTCGCATCAACCAACCGTTATT 3992
 Db 5129 TTCAACAGCGCAGCATTAACGCTGTCATCAAAATCACTCGCATCAACCAACCGTTATT 5070
 QY 3993 CATTCGTGATTGCGCTGAGCGAGACGAAATACGCGATCGCTGTTTAAAGGACAAATTACA 4052
 Db 5069 CATTCGTGATTGCGCTGAGCGAGACGAAATACGCGATCGCTGTTTAAAGGACAAATTACA 5010
 QY 4053 AACAGGAATCAATGCAACCGCGGAGAAACACATGCGCAGCGCATCAACATATTTTCCACC 4112
 Db 5009 AACAGGAATCAATGCAACCGCGGAGAAACACATGCGCAGCGCATCAACATATTTTCCACC 4950
 QY 4113 TGAATCAGGATATTTCTTAATACCTGGAATGCTTTTCCCGGGATCGCAGTGTGAG 4172
 Db 4949 TGAATCAGGATATTTCTTAATACCTGGAATGCTTTTCCCGGGATCGCAGTGTGAG 4890
 QY 4173 TAACCATGATCATCAGGAGTACGGAATAAATGCTTGAATGCTGGAAGAGGCAATAAATTC 4232
 Db 4889 TAACCATGATCATCAGGAGTACGGAATAAATGCTTGAATGCTGGAAGAGGCAATAAATTC 4830
 QY 4233 CGTCAGCGAGTTTATGCTGACCATCTCATCTGTAAACATCATTTGGCAACGCTACCTTTGCC 4292
 Db 4829 CGTCAGCGAGTTTATGCTGACCATCTCATCTGTAAACATCATTTGGCAACGCTACCTTTGCC 4770
 QY 4293 ATGTTTTCAGAAACAACTCTGCGCATCGGGCTTCCCATCAATCGATAGATTGCGCAC 4352
 Db 4769 ATGTTTTCAGAAACAACTCTGCGCATCGGGCTTCCCATCAATCGATAGATTGCGCAC 4710

QY 4353 TGATTGCCGACATTATCGGAGGCCATTTATACCATATATAAAATCAGCATCCATGTTGGA 4412
DB |||||||
QY 4709 TGATTGCCGACATTATCGGAGGCCATTTATACCATATATAAAATCAGCATCCATGTTGGA 4650
DB |||||||
QY 4413 ATTTAAATCGCGGCTCGAGCAAGACGTTTCCCGTTGAATATGCTCATACACCCCTTGT 4472
DB |||||||
QY 4649 ATTTAAATCGCGGCTCGAGCAAGACGTTTCCCGTTGAATATGCTCATACACCCCTTGT 4590
DB |||||||
QY 4473 ATTTACTGTTATGTAAGCAGACAGTTTATTTGTTATGTTATGTTATGTTATGTTGTC 4532
DB |||||||
QY 4589 ATTTACTGTTATGTAAGCAGACAGTTTATTTGTTATGTTATGTTATGTTATGTTGTC 4530
DB |||||||
QY 4533 AATGTACATCAGAGATTTTTCGAGACACAACTGCTTACTAGGATCCGATATCATTTAAA 4592
DB |||||||
QY 4529 AATGTACATCAGAGATTTTTCGAGACACAACTGCTTACTAGGATCCGATATCATTTAAA 4476
DB |||||||
QY 4593 TCTAGGATACAGGTAATACTAGTGTGCAACCAACAGATAAGTGAATCTAGTTCCAA 4652
DB |||||||
QY 4475 -----GGTGCACCAACAGATAAGTGAATCTAGTTCCAA 4441
DB |||||||
QY 4653 ACTATTTTGTGCTTTTAAATTTTCTGATATGCTTACGACGCTACACCCAGTTCCATCTA 4712
DB |||||||
QY 4440 ACTATTTTGTGCTTTTAAATTTTCTGATATGCTTACGACGCTACACCCAGTTCCATCTA 4381
DB |||||||
QY 4713 TTTTGTCACTCTCCCTAAATAATCTTAAATACTCCATTTTCCACCCCTCCAGTTCCCA 4772
DB |||||||
QY 4380 TTTTGTCACTCTCCCTAAATAATCTTAAATACTCCATTTTCCACCCCTCCAGTTCCCA 4321
DB |||||||
QY 4773 ACTATTTTCTACTCACTCAC 4792
DB |||||||
QY 4320 ACTATTTTGTGCGCCACAC 4301
DB |||||||

RESULT 10
ADG46818/C
ID ADG46818 standard; DNA; 5926 Bp.
XX
AC ADG46818;
XX
DT 11-MAR-2004 (first entry)
XX
DE PEM DNA (pBR plasmid containing a kanamycin mTn7 element).
XX
KW Transposon; ATP-utilising regulatory protein; gene; genetic analysis;
KW insertional mutagenesis; gene; plasmid; cyclic; circular; ds.
XX
OS Unidentified.
XX
PN US2002188105-A1.
XX
PD 12-DEC-2002.
XX
XX 19-DEC-2001; 2001US-00024809.
XX
XX 20-FEB-1997; 97US-0037955P.
XX
XX 20-FEB-1998; 98US-00027165.
XX
XX (UWJO) UNIV JOHNS HOPKINS SCHOOL MEDICINE.
XX
XX Craig NL;
XX
XX WPI; 2004-020185/02.
XX
XX Transposon for use in e.g. DNA sequencing methods, comprises a mutation
PT that allows efficient and simple insertion of and reduced target site
PT specificity on the transposon.
XX
XX Example 2; SEQ ID NO 4; 80pp; English.
PS
XX The invention relates to a transposon encoding an ATP-utilising
CC regulatory protein containing a mutation that allows efficient and simple
CC insertion of and reduced target site specificity on the transposon. The
CC invention is used in DNA sequencing methods, for genetic analysis by

CC insertional mutagenesis, and alteration of gene expression. The invention
CC is used to achieve efficient, simple, non-specific or random insertion
CC into any given DNA segment. The present sequence is pBR plasmid
CC containing a kanamycin mTn7 element) used in the exemplification of the
CC invention.
XX
SQ Sequence 5926 Bp; 1475 A; 1471 C; 1464 G; 1516 T; 0 U; 0 Other;
Query Match 30.6%; Score 1474; DB 12; Length 5926;
Best Local Similarity 93.3%; Pred. No. 9.7e-267;
Matches 1592; Conservative 0; Mismatches 40; Indels 74; Gaps 2;
QY 3087 TCCGTATGAGTGAAGTAAAGTCTTAACTGAACAAATAGATCTTAACTATGACAA 3146
DB |||||||
QY 5568 TCTAAATGTGGCGGACATTAAGTCTTAACTGAACAAATAGATCTTAACTATGACAA 5509
DB |||||||
QY 3147 TAAAGTCTTAACTAGACAGATAGTTGTAACCTGAATCACTCAGTCACTTATGCTGTGAAA 3206
DB |||||||
QY 5508 TAAAGTCTTAACTAGACAGATAGTTGTAACCTGAATCACTCAGTCACTTATGCTGTGAAA 5449
DB |||||||
QY 3207 AAGCATATCGGACTTTTGTATGCTTAAAGCAAACTCTTCAATTTCTGAAGTGCAAAATG 3266
DB |||||||
QY 5448 AAGCATATCGGACTTTTGTATGCTTAAAGCAAACTCTTCAATTTCTGAAGTGCAAAATG 5389
DB |||||||
QY 3267 CCCGTCTGTTTAAAGAGGGCGTGGGGTTCGACGGCGCTTAACTATATAACCGTCTCTAAGG 3326
DB |||||||
QY 5388 CCCGTCTGTTTAAAGAGGGCGTGGGGTTCGAC----- 5357
DB |||||||
QY 3327 TAGCGAGTTTAAACGATATCGGATCCGGCCGCGCTGAGGTCGCTCGTCAAGAGAGTG 3386
DB |||||||
QY 5356 -----CTGCAGGGGGGGGGGGCGCTGAGGTCGCTCGTCAAGAGAGTG 5312
DB |||||||
QY 3387 TTGCTGACTCATACAGGCTGAAATCGCCCATCATCCAGCAGAAAAGTGAAGGAGCCAC 3446
DB |||||||
QY 5311 TTGCTGACTCATACAGGCTGAAATCGCCCATCATCCAGCAGAAAAGTGAAGGAGCCAC 5252
DB |||||||
QY 3447 GGTGATGAGAGCTTTTGTGTAGGTGGAACCAAGTTGGTGATTTGAACTTTTGTGTTGCA 3506
DB |||||||
QY 5251 GGTGATGAGAGCTTTTGTGTAGGTGGAACCAAGTTGGTGATTTGAACTTTTGTGTTGCA 5192
DB |||||||
QY 3507 CGGAAACGCTGCTGCTGCTGGGGAAGATCGGTGATCTGATCTTCAACTCAGCAGAGATTC 3566
DB |||||||
QY 5191 CGGAAACGCTGCTGCTGCTGGGGAAGATCGGTGATCTGATCTTCAACTCAGCAGAGATTC 5132
DB |||||||
QY 3567 GATTTATTCACAAAGCGCGCTCCGTCAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 3626
DB |||||||
QY 5131 GATTTATTCACAAAGCGCGCTCCGTCAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 5072
DB |||||||
QY 3627 CCAATTAACCAATTCGATTTAGAAAAAATCTCATCGAGCATCAAAATCAAACTGCAATTTAT 3686
DB |||||||
QY 5071 CCAATTAACCAATTCGATTTAGAAAAAATCTCATCGAGCATCAAAATCAAACTGCAATTTAT 5012
DB |||||||
QY 3687 CATATCAGGATTTCAATACCATATTTTGAAGAACCGGTTTCTGTATGAGGAGGAGAAA 3746
DB |||||||
QY 5011 CATATCAGGATTTCAATACCATATTTTGAAGAACCGGTTTCTGTATGAGGAGGAGAAA 4952
DB |||||||
QY 3747 CTCACCGAGGAGTTTCCATAGGATCGGCAAGATCTCGGTATCGGTATCGGTATCGGTATCG 3806
DB |||||||
QY 4951 CTCACCGAGGAGTTTCCATAGGATCGGCAAGATCTCGGTATCGGTATCGGTATCGGTATCG 4892
DB |||||||
QY 3807 TCCAAATCAATACCAACCTTATTAATTTCCCTCGTCAAAAAATAGGTTATCAAGTGAGAA 3866
DB |||||||
QY 4891 TCCAAATCAATACCAACCTTATTAATTTCCCTCGTCAAAAAATAGGTTATCAAGTGAGAA 4832
DB |||||||
QY 3867 ATCACCATGAGTACGATGATCGGTGAGATCGGCAAGATCTCGGTATCGGTATCGGTATCG 3926
DB |||||||
QY 4831 ATCACCATGAGTACGATGATCGGTGAGATCGGCAAGATCTCGGTATCGGTATCGGTATCG 4772
DB |||||||
QY 3927 GACTTGTTCACAGCGCCGATTAAGTCTCGTCATCAAAATCACTCGCATCAACCAAAAC 3986
DB |||||||
QY 4771 GACTTGTTCACAGCGCCGATTAAGTCTCGTCATCAAAATCACTCGCATCAACCAAAAC 4712
DB |||||||
QY 3987 GTTATTCATTCGTGATTCGCTCGTTCGAGCGAGACAAATACGCGATCGCTGTTAAAGGACA 4046

| | | | |
|----|------|--|------|
| Db | 4711 | TTCA TTTCA TTTCTG GATTCGGCTCGAGCGAGACGAAATACGCGATCGCTGTTTAAAGACA | 4652 |
| Qy | 4047 | ATTACAAACAGGAATCGAAATCGAACCGCGCAGGAACTGCCAGCGCATCAACAATATT | 4106 |
| Db | 4651 | ATTACAAACAGGAATCGAATCGAACCGCGCAGGAACTGCCAGCGCATCAACAATATT | 4592 |
| Qy | 4107 | TTCACTTGAATCAGGATATCTTCTTAATACCTCGGAATGCTGTTTTCCGGGGATCGCAGT | 4166 |
| Db | 4591 | TTCACTTGAATCAGGATATCTTCTTAATACCTCGGAATGCTGTTTTCCGGGGATCGCAGT | 4532 |
| Qy | 4167 | GGTCAGTAACCATGCATCATCAGAGTACGAGTAACGATAAATGCTTGATGTCGGAAGGCAT | 4226 |
| Db | 4531 | GGTCAGTAACCATGCATCATCAGAGTACGAGTAACGATAAATGCTTGATGTCGGAAGGCAT | 4472 |
| Qy | 4227 | AAATTCCTGTCAGCCAGTTTACTGTCACCATCTCATCTGTAACATCATTTGGCAACGCTPACC | 4286 |
| Db | 4471 | AAATTCCTGTCAGCCAGTTTACTGTCACCATCTCATCTGTAACATCATTTGGCAACGCTPACC | 4412 |
| Qy | 4287 | TTTGCCATGTTTCAGAAACAACTCTGGCGCATCGGGCTTCCCATACAATCGATAGATTGT | 4346 |
| Db | 4411 | TTTGCCATGTTTCAGAAACAACTCTGGCGCATCGGGCTTCCCATACAATCGATAGATTGT | 4352 |
| Qy | 4347 | CGACCTGATTTGCCCGACATTTATCGCGAGGCCATTTTATACCCATATAAATCAGCATCCAT | 4406 |
| Db | 4351 | CGACCTGATTTGCCCGACATTTATCGCGAGGCCATTTTATACCCATATAAATCAGCATCCAT | 4292 |
| Qy | 4407 | GTTCGAAATTTAAATCGCGCCCTCGAGCAAGAGCTTTCCGGTTGAATATGGCTCATACACC | 4466 |
| Db | 4291 | GTTCGAAATTTAAATCGCGCCCTCGAGCAAGAGCTTTCCGGTTGAATATGGCTCATACACC | 4232 |
| Qy | 4467 | CCTTGTATTACTGTTTATGTAAGCAGACAGTTTTTATTTGTTTCATGATGATATATTTTATC | 4526 |
| Db | 4231 | CCTTGTATTACTGTTTATGTAAGCAGACAGTTTTTATTTGTTTCATGATGATATATTTTATC | 4172 |
| Qy | 4527 | TTTGCAATGTAAATCAGAGATTTTCAGACACAAACGCTGCTTACTAGGATCCGATATCA | 4586 |
| Db | 4171 | TTTGCAATGTAAATCAGAGATTTTCAGACACAAACGCTGCTTACTAGGATCCGATATCA | 4112 |
| Qy | 4587 | TTTAAATCTAGGATACAGGGTAATACTAGTGTCCGACCAACCCAGATAAGTGAATCTAG | 4646 |
| Db | 4111 | -----GGTCGACCAACAGATAAGTGAATCTAG | 4083 |
| Qy | 4647 | TTCCAAACTATTTTGTCAATTTTAAATTTTCGTAATTAGCTTTACGAGCGTCACACCAGTTCC | 4706 |
| Db | 4082 | TTCCAAACTATTTTGTCAATTTTAAATTTTCGTAATTAGCTTTACGAGCGTCACACCAGTTCC | 4023 |
| Qy | 4707 | CATCTATTTTGTCACTCTTCCCTAAATAATCTTTAAAAATCCATTTTCCACCCCTCCCGAG | 4766 |
| Db | 4022 | CATCTATTTTGTCACTCTTCCCTAAATAATCTTTAAAAATCCATTTTCCACCCCTCCCGAG | 3963 |
| Qy | 4767 | TTCCCAACTATTTTCTACTCACTCAC | 4792 |
| Db | 3962 | TTCCCAACTATTTTGTCCGCCACAC | 3937 |

RESULT 11

REGONI I-
AAF59457

ID AAF59457 standard; DNA; 4840 BP.

XX

AC AAF59457;

XX

DT 03-MAY-2001 (first entry)

XX

DE Npro-hGH expression plasmid pNPH1 nucleotide sequence.

XX

KW Human

KW auto

X

OS

OS Class

XX
"200"

PN WO200111057-A1.

[illegible]

Db 421 CACAGGTGCGTTCCTGGCGCTATATCGCGCATACCCGATCGGGAAGATCGGCTCG 480
Qy 1808 CCACCTCGGGCTCATGAGCGCTTCTTTCGGCGTGGGTATGTTGCGAGCCCGCTGGCGG 1867
Db 481 CCACTTCGGGCTCATGAGCGCTTCTTTCGGCGTGGGTATGTTGCGAGCCCGCTGGCGG 540
Qy 1868 GGGACTGTTGGCGCCATCTCTTGCATGACCACTTCTTTCGGCGGCGGTGCTCAACGG 1927
Db 541 GGGACTGTTGGCGCCATCTCTTGCATGACCACTTCTTTCGGCGGCGGTGCTCAACGG 600
Qy 1928 CCTCAACTACTACTGGGCTCTTCTTAATCAGAGAGTCGATAGGAGAGCGCTGCACC 1987
Db 601 CCTCAACTACTACTGGGCTCTTCTTAATCAGAGAGTCGATAGGAGAGCGCTGCACC 660
Qy 1988 GATGCCCTTGAGAGCTTCAACCCAGTCAGCTCTTTCGGTGGGCGGCGGCGACTAT 2047
Db 661 GATGCCCTTGAGAGCTTCAACCCAGTCAGCTCTTTCGGTGGGCGGCGGCGACTAT 720
Qy 2048 CGTCCGCGCACTATGACTGCTTCTTTATCATGCACTCGTAGGACAGGTGCCGCGAGC 2107
Db 721 CGTCCGCGCACTATGACTGCTTCTTTATCATGCACTCGTAGGACAGGTGCCGCGAGC 780
Qy 2108 GCTCTGGGTCACTTTCGGCGAGGACCGCTTTCGCTGAGCGCGACGATGATCGGCTGTC 2167
Db 781 GCTCTGGGTCACTTTCGGCGAGGACCGCTTTCGCTGAGCGCGACGATGATCGGCTGTC 840
Qy 2168 GCTTCGGGTATTCGGAATCTTTCGACGCGCTCGCTCAAGCCTTGTCTACTGTCGCGCAC 2227
Db 841 GCTTCGGGTATTCGGAATCTTTCGACGCGCTCGCTCAAGCCTTGTCTACTGTCGCGCAC 900
Qy 2228 CAAAGCTTCGGCGAGAGAGCGCACTTATCGCGGATGCGCGCGAGCGCTGGGCTA 2287
Db 901 CAAAGCTTCGGCGAGAGAGCGCACTTATCGCGGATGCGCGCGAGCGCTGGGCTA 960
Qy 2288 CGTCTGCTGGCTTCGGCGAGCGAGGCTTATGAGGCTTCCCATATGATTCCTCTCdc 2347
Db 961 CGTCTGCTGGCTTCGGCGAGCGAGGCTTATGAGGCTTCCCATATGATTCCTCTCdc 1020
Qy 2348 TTCGGCGGCGATCGGATGCGCGCTTTCAGGCGCATGCTGTCCAGGCGAGGTAGATGACda 2407
Db 1021 TTCGGCGGCGATCGGATGCGCGCTTTCAGGCGCATGCTGTCCAGGCGAGGTAGATGACda 1080
Qy 2408 CCATCAGGGAAGCTTCAAGGATCGCTCGCGCTCTTACCGCCTAACTTCGATCATtdg 2467
Db 1081 CCATCAGGGAAGCTTCAAGGATCGCTCGCGCTCTTACCGCCTAACTTCGATCATtdg 1140
Qy 2468 ACCGCTGATCGTCAGGCGATTTATGCGGCTCGCGGACATGGAACGGGTGGCATg 2527
Db 1141 ACCGCTGATCGTCAGGCGATTTATGCGGCTCGCGGACATGGAACGGGTGGCATg 1200
Qy 2528 GATTGAGCGCGCGCCCTATACCTTGTCTGCTCCCGGTTGCGTGGTATGAG 2587
Db 1201 GATTGAGCGCGCGCCCTATACCTTGTCTGCTCCCGGTTGCGTGGTATGAG 1260
Qy 2588 CCGGCGCACTCGACTGAATGGAAGCGCGGCGACCTCGCTAAAGGATTCACCACTCGA 2647
Db 1261 CCGGCGCACTCGACTGAATGGAAGCGCGGCGACCTCGCTAAAGGATTCACCACTCGA 1320
Qy 2648 AGAATGGAGCAATCAATTTCTTGGGAGAACTGTGAATGCGCAACCAACCTTTGGCAG 2707
Db 1321 AGAATGGAGCAATCAATTTCTTGGGAGAACTGTGAATGCGCAACCAACCTTTGGCAG 1380
Qy 2708 AACATATCATCGCTTCGCCATCTCCAGCAGCGCGCATCTTCGGGCGAGCGT 2767
Db 1381 AACATATCATCGCTTCGCCATCTCCAGCAGCGCGCATCTTCGGGCGAGCGT 1440
Qy 2768 GGGTCTTGGC 2778
Db 1441 GGGTCTTGGC 1451

AAQ29146/c
ID AAQ29146 standard; DNA; 3474 BP.
XX
AC AAQ29146;
XX
DT 24-OCT-2003 (revised)
DT 25-MAR-2003 (revised)
DT 05-MAR-1993 (first entry)
XX
DE pmTNF-MPH plasmid for expression of p362 from M. paratuberculosis.
XX
KW Mycobacterium; Crohn's disease; Johne's disease; cattle; human; ss.
XX
OS Mycobacterium avium subsp. paratuberculosis.
XX
FH Key Location/Qualifiers
FT misc_feature 1..208
FT /tag= a
FT /note= "lambda PL contg. EcoRI blunt-MboI blunt fragment
FT of pPL(lambda) "
FT 209..436
FT /tag= b
FT /note= "synthetic DNA fragment"
FT 230..307
FT /tag= c
FT /note= "encodes residues 2-25 of mature mouse TNF"
FT 308..384
FT /tag= d
FT /note= "multiple cloning site contg. 6 His encoding
FT sequence at position 315-332"
FT 385..436
FT /tag= e
FT /note= "HindIII fragment contg. E. coli trp terminator"
FT 437..943
FT /tag= f
FT /note= "rrnBT12 contg. HindIII-SapI fragment from
FT pKK223"
FT 944..3474
FT /tag= g
FT /note= "DraI-EcoRI blunt fragment of pAT153 contg.
FT tetracycline resistance gene and origin of replication"
XX
PN W09216628-A1.
PD 01-OCT-1992.
XX
PF 24-MAR-1992; 92WO-EP0000661.
XX
PR 25-MAR-1991; 91EP-00400798.
XX
PA (INNO-) INNOGENETICS NV SA.
XX
PI Cocito C, Coene M, De Kesel M, Gilot P;
XX WPI; 1992-349213/42.
XX
PT Polypeptide(s) derived from mycobacterium para-tuberculosis - are
PT immunogens, useful as vaccines and for diagnosing Crohns disease.
XX
PS Disclosure; Fig 9; 105pp; English.
XX
CC The DNA sequence is that of the expression vector pmTNF-MPH which
CC contains the tetracycline resistance gene and origin of replication of
CC pAT153, the lambda PL promoter up to the MboI site in the N gene 5'
CC untranslated region, followed by a synthetic ribosome binding site and
CC the information encoding the first 25 amino acids of mouse TNF. This
CC sequence is followed by a synthetic polylinker encoding six consecutive
CC histidines followed by several proteolytic sites each accessible via a
CC different restriction enzyme which is unique for the plasmid. Downstream
CC from the polylinker, several transcription terminators are present, of the
CC including the E. coli trp terminator. The DNA encoding a fragment of the
CC 34 kD p362 protein from Mycobacterium paratuberculosis was subcloned into
CC this expression vector to express the fusion protein mTNF-H6-p362. The

| | | |
|----|---|--|
| PT | nucleic acid, antibodies, anti-sense cpds. etc. | |
| XX | Disclosure; Fig 8b; 108pp; English. | |
| PS | | |
| XX | The plasmid is used to clone and express. The polypeptide induced in | |
| CC | macrophages by lipopolysaccharide stimulates cell proliferation (esp. | |
| CC | when costimulated with IL-4) promote activation, cytotoxicity, and | |
| CC | mobilisation of LAK cells; promote recruitment of suppressive peritoneal | |
| CC | exudate cells; promote generation of immunocompetent lymph node cells | |
| CC | (LNC) and have trypanocidal and trypanolytic activity. The human and | |
| CC | murine sequences are given in (AA051543-45), peptide fragments able to | |
| CC | generate antibodies are given in (AAR51951-61) (Updated on 25-MAR-2003 to | |
| CC | correct PN field.) | |
| XX | | |
| SQ | Sequence 3474 BP; 845 A; 933 C; 978 G; 718 T; 0 U; 0 Other; | |
| | Query Match 30.1%; Score 1446.8; DB 2; Length 3474; | |
| | Best Local Similarity 99.9%; Pred. No. 1.2e-261; | |
| | Matches 1448; Conservative 0; Mismatches 2; Indels 0; Gaps 0; | |
| Qy | 1329 AATTCTCATGTTGACAGCTTATCATGATAAGCTTTAATGCCGTAGTTATACAGTTA 1388 | |
| Db | 3474 AATTCTCATGTTGACAGCTTATCATGATAAGCTTTAATGCCGTAGTTATACAGTTA 3415 | |
| Qy | 1389 AATTGCTAAGCAGTFCAGGCACCGTGTATGAATCTTAACAATGCGTCAATCGTCATCTC 1448 | |
| Db | 3414 AATTGCTAAGCAGTFCAGGCACCGTGTATGAATCTTAACAATGCGTCAATCGTCATCTC 3355 | |
| Qy | 1449 GGCACCGTCAACCTGGATGCTGTAGGCATAGGCTTGGTTATGCCGGTACTGCCGGGCTC 1508 | |
| Db | 3354 GGCACCGTCAACCTGGATGCTGTAGGCATAGGCTTGGTTATGCCGGTACTGCCGGGCTC 3295 | |
| Qy | 1509 TTGGGGGATATCGTCCATTCGACAGCATCCGACGTCACTATGCGGTGCTGTACGCGTA 1568 | |
| Db | 3294 TTGGGGGATATCGTCCATTCGACAGCATCCGACGTCACTATGCGGTGCTGTACGCGTA 3235 | |
| Qy | 1569 TATCGTTGATGCAATTTCTATGCGCACCGTTCTCGGAGCAGTTCGCGCGCTTTGGC 1628 | |
| Db | 3234 TATCGTTGATGCAATTTCTATGCGCACCGTTCTCGGAGCAGTTCGCGCGCTTTGGC 3175 | |
| Qy | 1629 CGCGGCCAGTCTGCTCGCTTCGCTACTTTGGAGCCACTATCGACTACGCGCATATGGCG 1688 | |
| Db | 3174 CGCGGCCAGTCTGCTCGCTTCGCTACTTTGGAGCCACTATCGACTACGCGCATATGGCG 3115 | |
| Qy | 1689 ACCACACCCGCTCTGTGTGATCTCTACGCCGAGCAGCATGTGGCCGGCATCACCGGCGC 1748 | |
| Db | 3114 ACCACACCCGCTCTGTGTGATCTCTACGCCGAGCAGCATGTGGCCGGCATCACCGGCGC 3055 | |
| Qy | 1749 ACAGTGGGTTGCTGGCGCCTATATCGCGACATCACCGATGGGGAAGATCGGCTCGC 1808 | |
| Db | 3054 ACAGTGGGTTGCTGGCGCCTATATCGCGACATCACCGATGGGGAAGATCGGCTCGC 2995 | |
| Qy | 1809 CACTTCGGGGCTCATGAGCGCTTGTTCGGCGTGGGTATGTTGGCAGGCCCGCTGGCCGG 1868 | |
| Db | 2994 CACTTCGGGGCTCATGAGCGCTTGTTCGGCGTGGGTATGTTGGCAGGCCCGCTGGCCGG 2935 | |
| Qy | 1869 GAGCTGTTGGCGCCATCTCTTGATGACACCATTCCTTGGCGCGCGCGTGTCAACGCG 1928 | |
| Db | 2934 GAGCTGTTGGCGCCATCTCTTGATGACACCATTCCTTGGCGCGCGCGTGTCAACGCG 2875 | |
| Qy | 1929 CTCACCTACTACTGGGTGCTTCTTAATGAGGAGTGCATTAAGGGAGAGGCTGCACCG 1988 | |
| Db | 2874 CTCACCTACTACTGGGTGCTTCTTAATGAGGAGTGCATTAAGGGAGAGGCTGCACCG 2815 | |
| Qy | 1989 ATGCCCTTGAGAGCTTCAACCCAGTCAGTCTCTTCGGTGGCGCGGGGATGACTATC 2048 | |
| Db | 2814 ATGCCCTTGAGAGCTTCAACCCAGTCAGTCTCTTCGGTGGCGCGGGGATGACTATC 2755 | |
| Qy | 2049 GTGCCGCGCATTTATGACTGTCTTCTTTATCATGCAACTCGTAGGACAGGTGCCGCGCG 2108 | |
| Db | 2754 GTGCCGCGCATTTATGACTGTCTTCTTTATCATGCAACTCGTAGGACAGGTGCCGCGCG 2695 | |
| Qy | 2109 CTCTGGGTCAATTTTCGGCGAGGACCGCTTTCGCTGGAGCGGACGATGATCGGCCGTGCG 2168 | |

| | | |
|----|--|--|
| Db | 2694 CTCTGGGTCAATTTTCGGCGAGGACCGCTTTCGTTGGAGCGGACGATGATCGGCTGTGCG 2635 | |
| Qy | 2169 CTTGCGGTATTTTCGGAATCTTTGACGCGCTCGCTCAAGCCTTCGCTCACTGTGTCGCCACAC 2228 | |
| Db | 2634 CTTGCGGTATTTTCGGAATCTTTGACGCGCTCGCTCAAGCCTTCGCTCACTGTGTCGCCACAC 2575 | |
| Qy | 2229 AAACGTTTTCGGCGAGAGCAGGCGCATTTATCCCGGCGATGCGGCGGACGCGCTGGGTAC 2288 | |
| Db | 2574 AAACGTTTTCGGCGAGAGCAGGCGCATTTATCCCGGCGATGCGGCGGACGCGCTGGGTAC 2515 | |
| Qy | 2289 GTCTTGTGCGGTTTCGCGAGCGGAGGCTGATGCGCCTTCCCATATGATTTCTTCGCT 2348 | |
| Db | 2514 GTCTTGTGCGGTTTCGCGAGCGGAGGCTGATGCGCCTTCCCATATGATTTCTTCGCT 2455 | |
| Qy | 2349 TCCGGCGGCATCGGGATGCGCGGTTTCGAGGCCATGCTGTCCAGGAGGTAGATGACGAC 2408 | |
| Db | 2454 TCCGGCGGCATCGGGATGCGCGGTTTCGAGGCCATGCTGTCCAGGAGGTAGATGACGAC 2395 | |
| Qy | 2409 CATCAGGACACAGCTTCAAGGATCGCTCGCGCTCTTACAGCCTTAACCTTCGATCAATTGGA 2468 | |
| Db | 2394 CATCAGGACACAGCTTCAAGGATCGCTCGCGCTCTTACAGCCTTAACCTTCGATCAATTGGA 2335 | |
| Qy | 2469 CCGCTGATCTGTCACGGCGATTTATGCGCGCTTCGCGGAGCAGCATGGAACGGGTTGGCATGG 2528 | |
| Db | 2334 CCGCTGATCTGTCACGGCGATTTATGCGCGCTTCGCGGAGCAGCATGGAACGGGTTGGCATGG 2275 | |
| Qy | 2529 ATTGTAGCGCGCGCTTACCTTGTCTGCTCTCCCGGTTGCGTCCGCGTGCATGGAGC 2588 | |
| Db | 2274 ATTGTAGCGCGCGCTTACCTTGTCTGCTCTCCCGGTTGCGTCCGCGTGCATGGAGC 2215 | |
| Qy | 2589 CGGCGCACCTCGACCTCAATGGAAGCGCGGCGCACCTCGCTTAACGGATTCACCACTCCAA 2648 | |
| Db | 2214 CGGCGCACCTCGACCTCAATGGAAGCGCGGCGCACCTCGCTTAACCGATTCACCACTCCAA 2155 | |
| Qy | 2649 GAATTGGAGCCAAATCAATTTCTTTCGCGAGAACTGTGAATCGCGCAAAACCAACCTTGGCAGA 2708 | |
| Db | 2154 GAATTGGAGCCAAATCAATTTCTTTCGCGAGAACTGTGAATCGCGCAAAACCAACCTTGGCAGA 2095 | |
| Qy | 2709 ACATATCCATCGCTCGCCCATCTCCAGAGCCCGCGCGCATCTCGCGGCGAGCGTTG 2768 | |
| Db | 2094 ACATATCCATCGCTCGCCCATCTCCAGAGCCCGCGCGCATCTCGCGGCGAGCGTTG 2035 | |
| Qy | 2769 GGTCTCTGGCG 2778 | |
| Db | 2034 GGTCTCTGGCC 2025 | |
| | RESULT 14 | |
| | ADL72227/c | |
| ID | ADL72227 standard; DNA; 3801 BP. | |
| XX | ADL72227; | |
| AC | AC | |
| XX | 01-JUL-2004 (first entry) | |
| DT | 01-JUL-2004 (first entry) | |
| XX | DNA sequence of plasmid construct, pACSE. | |
| DE | Evolutionary potential; mutant resistance gene; anti-pathogenic drug; | |
| XX | pathogen; ds. | |
| KW | Synthetic. | |
| XX | OS | |
| XX | US6720142-B1. | |
| PN | 13-APR-2004. | |
| XX | 18-AUG-2000; 2000US-00640882. | |
| PF | 19-AUG-1999; 99US-0149813P. | |
| XX | (UVRP) UNIV ROCHESTER. | |
| XX | | |

KW Vector; pACYC184; origin of replication; p15A; ds.
XX Escherichia coli.
XX WO2003008556-A2.
XX 30-JAN-2003.
XX 18-JUL-2002; 2002WO-US023089.
XX 18-JUL-2001; 2001US-0306344P.
XX (EXPR-) EXPRESSIVE CONSTRUCTS INC.
XX Sanders MC, Hamilton M;
XX WPI; 2003-239326/23.
XX New p15A origin of replication, useful in expressing and/or co-expressing
XX proteins, in purifying a cloned nucleic acid, and in converting a low
XX copy number plasmid into a more efficient high copy number plasmid.
XX Disclosure; Page 48-49; 53pp; English.
XX The present sequence is the DNA sequence of plasmid pACYC184. The plasmid
XX includes an origin of replication sequence, denoted p15A ori, positioned
XX within an EcoRI-HindIII fragment corresponding to nucleotides 1-1523 of
XX the present sequence. The present invention is based on the finding that
XX key mutations in the p15A ori sequence operatively linked within a
XX plasmid result in the plasmid being converted from a low copy number to a
XX high copy number plasmid useful for more efficient cloning, DNA and RNA
XX purification, protein expression and co-expression. The mutated
XX (enhanced) p15A ori can be isolated and inserted into low copy plasmids
XX other than pACYC184 to convert them to high copy number plasmids.
XX Preferred mutations to p15A ori were identified following random
XX mutagenesis of the native p15A locus of pACYC184 by PCR, use of an
XX amplified 1.5 kb product to construct a plasmid library, transformation
XX of Escherichia coli BL21(DE3) cells and screening of transformed colonies
XX for an operationally enhanced p15A ori. Plasmid pBci-073 was obtained,
XX which has a copy number in E. coli BL21(DE3) about 4.47-fold higher than
XX that of pACYC184. Sequencing of the EcoRI-HindIII fragment of pBci-073
XX revealed 10 base substitutions (see AB270621) compared with the native
XX sequence
XX
XX SQ Sequence 4245 BP; 956 A; 1167 C; 1094 G; 1028 T; 0 U; 0 Other;
Query Match 30.1%; Score 1446.8; DB 8; Length 4245;
Best Local Similarity 99.9%; Pred. No. 1.2e-261;
Matches 1448; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1329 AATTCTCATGTTTGACAGCTTATCATCGATAAGCTTTAATGCGGTAGTTTATCACAGTTA 1388
DB 1494 AATTCTCATGTTTGACAGCTTATCATCGATAAGCTTTAATGCGGTAGTTTATCACAGTTA 1553
QY 1389 AATTGCTTAACGACGTCAGGACCGGTATGAATCTACAAATGCGCTCATCGTATCTCTC 1448
DB 1554 AATTGCTTAACGACGTCAGGACCGGTATGAATCTACAAATGCGCTCATCGTATCTCTC 1613
QY 1449 GGCACCGTCAACCTCGATGCTGTAGGCATAGGCTTTGGTTATGCCGGTACTGCGGGGCTC 1508
DB 1614 GGCACCGTCAACCTCGATGCTGTAGGCATAGGCTTTGGTTATGCCGGTACTGCGGGGCTC 1673
QY 1509 TTGCGGGATATCGTCCATTCGACAGCATCGCCAGTCAATATGGGTGCTGTAGCGCTA 1568
DB 1674 TTGCGGGATATCGTCCATTCGACAGCATCGCCAGTCAATATGGGTGCTGTAGCGCTA 1733
QY 1569 TATGCGTGTATGCAATTTCTATGCGCACCCGTTCTCGGAGCACTGTCGACCGCTTTGCG 1628
DB 1734 TATGCGTGTATGCAATTTCTATGCGCACCCGTTCTCGGAGCACTGTCGACCGCTTTGCG 1793
QY 1629 GCGCGCCAGTCTGCTCGCTTCGCTACTTTGGAGCCACTATCGACTACGCGCATCATGCGG 1688
DB 1794 GCGCGCCAGTCTGCTCGCTTCGCTACTTTGGAGCCACTATCGACTACGCGCATCATGCGG 1853

QY 1689 ACCACACCCGTCCTGTGGATCCTCTACGCCGAGCGCATCGTGGCCGCGCATCACCGGGGCC 1748
DB 1854 ACCACACCCGTCCTGTGGATCCTCTACGCCGAGCGCATCGTGGCCGCGCATCACCGGGGCC 1913
QY 1749 ACAGGTGCGGTTGCTGGCGCTATATCGCCGACATCACCGATGGGGAGATCGGGGCTCGC 1808
DB 1914 ACAGGTGCGGTTGCTGGCGCTATATCGCCGACATCACCGATGGGGAGATCGGGGCTCGC 1973
QY 1809 CACTTGGGGCTCATGAGCGGCTTTTTCGGCGTGGGTATGGTGGCAGGCCCGCTGGCCGGG 1868
DB 1974 CACTTGGGGCTCATGAGCGGCTTTTTCGGCGTGGGTATGGTGGCAGGCCCGCTGGCCGGG 2033
QY 1869 GGACTGTTGGCGGCATCTCTCTTCATGTCACCAATTCCTTGGCGCGGGTCTCTCAACGGGC 1928
DB 2034 GGACTGTTGGCGGCATCTCTCTTCATGTCACCAATTCCTTGGCGCGGGTCTCTCAACGGGC 2093
QY 1929 CTCAACCTACTACTGGGCTGCTTCTTAATCAGGAGTCGATAGAGGAGAGCGTCGACCG 1988
DB 2094 CTCAACCTACTACTGGGCTGCTTCTTAATCAGGAGTCGATAGAGGAGAGCGTCGACCG 2153
QY 1989 ATGCCCTTGAGAGCTTCAACCCAGTCAGCTCCTTCGGTGGGCGCGGGCATGACTATC 2048
DB 2154 ATGCCCTTGAGAGCTTCAACCCAGTCAGCTCCTTCGGTGGGCGCGGGCATGACTATC 2213
QY 2049 GTCCGCCACCTTATGACTGCTCTTCTTATCATGCAACTCGTAGGACAGAGTCCCGGACGC 2108
DB 2214 GTCCGCCACCTTATGACTGCTCTTCTTATCATGCAACTCGTAGGACAGAGTCCCGGACGC 2273
QY 2109 CTCTGGGTCAATTTTCGGCGAGGACCGCTTCGCTGGAGCGGCGCATGATCGGCTGTGCG 2168
DB 2274 CTCTGGGTCAATTTTCGGCGAGGACCGCTTCGCTGGAGCGGCGCATGATCGGCTGTGCG 2333
QY 2169 CTTCGGGTATTCGGAATCTTGCAAGCCCTCGCTCAAGCCTTCGTCACCTGGTCCCGCAC 2228
DB 2334 CTTCGGGTATTCGGAATCTTGCAAGCCCTCGCTCAAGCCTTCGTCACCTGGTCCCGCAC 2393
QY 2229 AAAGCTTTCGCGAGAGACAGCGCATTAATCCCGCATGGCGCGCGAGCGCTGGGCTAC 2288
DB 2394 AAAGCTTTCGCGAGAGACAGCGCATTAATCCCGCATGGCGCGCGAGCGCTGGGCTAC 2453
QY 2289 GTCTTGTGCGGTTTCGCGACGCGAGGCTGGATGCGCTTCCCATTTATGATTTCTTCGCT 2348
DB 2454 GTCTTGTGCGGTTTCGCGACGCGAGGCTGGATGCGCTTCCCATTTATGATTTCTTCGCT 2513
QY 2349 TCCGCGCGCATCGGATGCGCGCTTCGAGCGCATGCTGTCCAGCGAGTATAGTACACAC 2408
DB 2514 TCCGCGCGCATCGGATGCGCGCTTCGAGCGCATGCTGTCCAGCGAGTATAGTACACAC 2573
QY 2409 CATCAGGACAGCTTCAAGGATCGCTCGCGCTCTTACAGCCTAACTTCGATCAATTGGA 2468
DB 2574 CATCAGGACAGCTTCAAGGATCGCTCGCGCTCTTACAGCCTAACTTCGATCAATTGGA 2633
QY 2469 CCGTGTATCGTACCGGCGATTTATGCGCGCTTCGCGAGCACATGGAACGGGTGGCATGG 2528
DB 2634 CCGTGTATCGTACCGGCGATTTATGCGCGCTTCGCGAGCACATGGAACGGGTGGCATGG 2693
QY 2529 ATTGTAGCGCGCGCTTATACCTTGTCTGCTCCCGGTTTCGCTCGCGGTGCTAGGAGC 2588
DB 2694 ATTGTAGCGCGCGCTTATACCTTGTCTGCTCCCGGTTTCGCTCGCGGTGCTAGGAGC 2753
QY 2589 CGGGCCACTCGACTGAATGAAGCGCGCGCACCTCGCTAAACGGATTCACCACTCCCAA 2648
DB 2754 CGGGCCACTCGACTGAATGAAGCGCGCGCACCTCGCTAAACGGATTCACCACTCCCAA 2813
QY 2649 GAAATGGAGCCCAATCAATTTCTTGGCGGAACTGTGAATGCGGAAACCAACCCCTTGGCAGA 2708
DB 2814 GAAATGGAGCCCAATCAATTTCTTGGCGGAACTGTGAATGCGGAAACCAACCCCTTGGCAGA 2873
QY 2709 ACATATTCATCGCTCGCGCATCTCCAGCAGCGCGGCGCATCTCGGCGACGCTG 2768
DB 2874 ACATATTCATCGCTCGCGCATCTCCAGCAGCGCGGCGCATCTCGGCGACGCTG 2933

Qy 2769 GGTCCTGGCC 2778
Db 2934 GGTCCTGGCC 2943

Search completed: January 17, 2006, 19:35:29
Job time : 3011.54 secs

GGTCCTGGCC 2778
GGTCCTGGCC 2943

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2006 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 17, 2006, 18:09:28 ; Search time 20375.3 Seconds
(without alignments)
11054.239 Million cell updates/sec

Title: US-10-511-327-7

Perfect score: 4814

Sequence: 1 ggtacctggaatgcgca.....taagatgttttctgtgact 4814

Scoring table:
IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST.*

- 1: gb_est1.*
- 2: gb_est2.*
- 3: gb_est3.*
- 4: gb_hc.*
- 5: gb_est4.*
- 6: gb_est5.*
- 7: gb_est6.*
- 8: gb_est7.*
- 9: gb_gss1.*
- 10: gb_gss2.*
- 11: gb_gss3.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------------|--------------------|
| C 1 | 920.4 | 19.1 | 933 | 9 B2430219 | B2430219 BOND117TF |
| C 2 | 918 | 19.1 | 918 | 9 B2430224 | B2430224 BOND117TR |
| C 3 | 801 | 16.6 | 881 | 9 AZ682163 | AZ682163 ENTW91TR |
| 4 | 719.2 | 14.9 | 988 | 11 CNS06X6V | AL419357 T7 end of |
| 5 | 704.8 | 14.6 | 889 | 6 CF577494 | CF577494 AGENCOURT |
| C 6 | 670.2 | 13.9 | 675 | 9 BH600610 | BH600610 BOHQ87TF |
| C 7 | 658 | 13.7 | 663 | 10 CG411076 | CG411076 RM391 Lxx |
| 8 | 648 | 13.5 | 648 | 1 AJ762384 | AJ762384 AJ762384 |
| 9 | 648 | 13.5 | 648 | 1 AJ762542 | AJ762542 AJ762542 |
| 10 | 646 | 13.4 | 659 | 2 BF632491 | BF632491 NF027F11D |
| 11 | 645 | 13.4 | 645 | 5 BQ752029 | BQ752029 EST632592 |
| C 12 | 643 | 13.4 | 681 | 10 CG804998 | CG804998 1118056B1 |
| 13 | 641 | 13.3 | 675 | 3 BJ684207 | BJ684207 BJ684207 |
| 14 | 641 | 13.3 | 676 | 3 BJ683711 | BJ683711 BJ683711 |
| 15 | 637.4 | 13.2 | 683 | 3 BJ683533 | BJ683533 BJ683533 |
| 16 | 636.4 | 13.2 | 675 | 3 BJ683290 | BJ683290 BJ683290 |
| 17 | 636 | 13.2 | 675 | 3 BJ683026 | BJ683026 BJ683026 |
| C 18 | 635.4 | 13.2 | 945 | 9 BH146743 | BH146743 ENTWJ23TF |
| 19 | 635 | 13.2 | 651 | 3 BJ684301 | BJ684301 BJ684301 |
| 20 | 635 | 13.2 | 670 | 3 BJ683987 | BJ683987 BJ683987 |
| C 21 | 628.4 | 13.1 | 947 | 9 AZ541608 | AZ541608 ENTWU74TF |
| 22 | 627.6 | 13.0 | 643 | 3 BJ682782 | BJ682782 BJ682782 |

| | | | | | |
|------|-------|------|------|-------------|--------------------|
| C 23 | 620.8 | 12.9 | 647 | 9 BH241665 | BH241665 ATYOB72TR |
| 24 | 617.8 | 12.8 | 650 | 3 BJ683212 | BJ683212 BJ683212 |
| 25 | 616 | 12.8 | 645 | 3 BJ682197 | BJ682197 BJ682197 |
| 26 | 613.2 | 12.7 | 658 | 3 BJ684166 | BJ684166 BJ684166 |
| 27 | 612.2 | 12.7 | 785 | 2 BE572157 | BE572157 601330463 |
| C 28 | 604.4 | 12.6 | 618 | 9 CC943909 | CC943909 BOIHE17TR |
| C 29 | 604 | 12.5 | 979 | 9 BH146752 | BH146752 ENTWJ23TR |
| C 30 | 591.8 | 12.3 | 599 | 9 CC964255 | CC964255 BOIEB66TF |
| 31 | 587.2 | 12.2 | 710 | 10 AG011497 | AG011497 Homo sapi |
| 32 | 584 | 12.1 | 638 | 1 AJ762641 | AJ762641 AJ762641 |
| C 33 | 582.6 | 12.1 | 1059 | 1 AL575254 | AL575254 AL575254 |
| 34 | 581.8 | 12.1 | 585 | 5 BU668685 | BU668685 MC01033H0 |
| C 35 | 572 | 11.9 | 572 | 9 CC948132 | CC948132 BOIEC75TF |
| C 36 | 566.2 | 11.8 | 579 | 5 BU719372 | BU719372 SJMCTE08 |
| C 37 | 565.2 | 11.7 | 570 | 9 BH241973 | BH241973 ATYQC33TF |
| C 38 | 561.8 | 11.7 | 580 | 9 CC967108 | CC967108 BOIHM43TF |
| C 39 | 560 | 11.6 | 574 | 7 CJ326964 | CJ326964 CJ326964 |
| C 40 | 556.6 | 11.6 | 745 | 9 BZ054578 | BZ054578 jnr34a02. |
| C 41 | 554.6 | 11.5 | 716 | 9 BZ050418 | BZ050418 jnr41b09. |
| 42 | 546.4 | 11.4 | 561 | 3 BJ682307 | BJ682307 BJ682307 |
| C 43 | 544.6 | 11.3 | 716 | 9 BZ050188 | BZ050188 jnr43b08. |
| C 44 | 543.6 | 11.3 | 715 | 9 BZ054871 | BZ054871 jnr32h05. |
| C 45 | 543.6 | 11.3 | 721 | 9 BZ054951 | BZ054951 jnr32h05. |

ALIGNMENTS

RESULT 1
BZ430219/c BZ430219 933 bp DNA linear GSS 13-DEC-2002
LOCUS BOND117TF BO.1.6.2_KB_tot Brassica oleracea genomic clone BOND117,
DEFINITION genomic survey sequence.
ACCESSION BZ430219
VERSION BZ430219.1 GI:26675100
KEYWORDS GSS
SOURCE Brassica oleracea
ORGANISM Brassica oleracea
REFERENCE 1 (bases 1 to 933)
AUTHORS Ayele,M., Haas,B.J., Kumar,N., Wu,H., Xiao,Y., Van Aken,S.,
Unterback,T.R., Wortman,J.R., White,O.R. and Town,C.D.
TITLE Whole genome shotgun sequencing of Brassica oleracea and its
application to gene discovery and annotation in Arabidopsis
JOURNAL Genome Res. 15 (4), 487-495 (2005)
PUBMED 15805490
COMMENT Other GSSs: BOND117TR
Contact: Chris Town
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TF
Class: sheared ends.
Location/Qualifiers
source
1. .933
/organism="Brassica oleracea"
/mol_type="genomic DNA"
/strain="T01000DH3"
/db_xref="taxon:3712"
/clone="BOND117"
/clone_lib="BO.1.6.2_KB_tot"
/note="vector: PHOS1; Site 1: BstXI; 1.6-2 kb sheared
total DNA inserted into PHOS1 using BstXI linkers"

ORIGIN

Query Match 19.1%; Score 920.4; DB 9; Length 933;
Best Local Similarity 99.8%; Pred. No. 2.1e-246;
Matches 932; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

```
Qy 1786 CCGATGGGGAAGATCGGGCTCGCCACTTCGGGCTCATGAGCGCTTGTTTTCGGCGTGGGTA 1845
Db 933 CCGATGGGGAAGATCGGGCTCGCCACTTCGGGCTCATGAGCGCTTGTTTTCGGCGTGGGTA 874
Qy 1846 TGGTGGCAGGCCCCGCTGGCCGGGGGACGTGTTGGGCGGCATCTCCTTGCATGCACCATTC 1905
Db 873 TGGTGGCAGGCCCCGCTGGCCGGGGGACGTGTTGGGGGCATCTCCTTGCATGCACCATTC 814
Qy 1906 TTGCGGCGGGGCTGCTCAACCGGCTCAACCTACTACTGCGGTGCTTCTTAATGAGGAGT 1965
Db 813 TTGCGGCGGGCTGCTCAACCGGCTCAACCTACTACTGCGGTGCTTCTTAATGAGGAGT 754
Qy 1966 CGCATAGGAGAGCGGTGACCGATGCCCTTGAGAGCCTTCAACCCAGTCACTCTTC 2025
Db 753 CGCATAGGAGAGCGGTGACCGATGCCCTTGAGAGCCTTCAACCCAGTCACTCTTC 694
Qy 2026 GGTGGGCGGGGATGACTATCGTTCGGCGCACTTATGACTGTCTTCTTTATCATGAAC 2085
Db 693 GGTGGGCGGGGATGACTATCGTTCGGCGCACTTATGACTGTCTTCTTTATCATGAAC 634
Qy 2086 TCGTAGGACAGGTGCGGACGCTCTGGGTCAATTTTCGGGAGAGACGCTTTCGCTGGA 2145
Db 633 TCGTAGGACAGGTGCGGACGCTCTGGGTCAATTTTCGGGAGAGACGCTTTCGCTGGA 574
Qy 2146 GCGGACGATCATCGGCTCTGCTTGGGTATTCGGATCTTGCACGCCCTGCTCAAG 2205
Db 573 GCGGACGATCATCGGCTCTGCTTGGGTATTCGGATCTTGCACGCCCTGCTCAAG 514
Qy 2206 CTTTCGTCACTGCTCCGCGACCAAAAGCTTTTCGGGAGAGAGCCCATATATCGCGGCA 2265
Db 513 CTTTCGTCACTGCTCCGCGACCAAAAGCTTTTCGGGAGAGAGCCCATATATCGCGGCA 454
Qy 2266 TGGCGGCGGACGCGCTGGGCTACGCTTGTGCGGTTCGCGACGCGAGGCTTGGATGGCT 2325
Db 453 TGGCGGCGGACGCGCTGGGCTACGCTTGTGCGGTTCGCGACGCGAGGCTTGGATGGCT 394
Qy 2326 TCCCATATATGATCTTCTGCTTCGGCGGATCGGATGCCCGGTGCGAGGCCATGC 2385
Db 393 TCCCATATATGATCTTCTGCTTCGGCGGATCGGATGCCCGGTGCGAGGCCATGC 334
Qy 2386 TGTCCAGCAGGTAGATCAGCACCATCAGGAGACGCTTCAAGGATCGCTCCGCGCTCTTA 2445
Db 333 TGTCCAGCAGGTAGATCAGCACCATCAGGAGACGCTTCAAGGATCGCTCCGCGCTCTTA 274
Qy 2446 CCAGCCTAACTTCGATCATTGGAACGCTGATCGTCAAGCGGATTTATGCCGCTCGGCGA 2505
Db 273 CCAGCCTAACTTCGATCATTGGAACGCTGATCGTCAAGCGGATTTATGCCGCTCGGCGA 214
Qy 2506 GCATGGAACGGGTGGATGGATTTGAGGCGCGCCCTATACCTTGTCTGCTTCCCG 2565
Db 213 GCATGGAACGGGTGGATGGATTTGAGGCGCGCCCTATACCTTGTCTGCTTCCCG 154
Qy 2566 GCTTCGCTCGGTCGATGGAGCGGGGACCTCGACCTGAATGGAAGCGGCGGACCT 2625
Db 153 GCTTCGCTCGGTCGATGGAGCGGGGACCTCGACCTGAATGGAAGCGGCGGACCT 94
Qy 2626 GCCTAACCGATTCACCATCCAAAGAAATGGAGCCAAATCAATTTCTTGGGAGAACTGTGAA 2685
Db 93 GCCTAACCGATTCACCATCCAAAGAAATGGAGCCAAATCAATTTCTTGGGAGAACTGTGAA 34
Qy 2686 TGGCAAAACCAACCTTGGGAGAAATATCCATC 2719
Db 33 TGGCAAAACCAACCTTGGGAGAAATATCCATC 1
```

RESULT 2
B2430224
LOCUS
DEFINITION
B2430224
ACCESSION
VERSION

BZ430224
BOND17TR BO.1.6.2_KB tot Brassica oleracea genomic clone BOND17,
genomic survey sequence.
BZ430224
BZ430224.1 GI:26675106

KEYWORDS

SOURCE

ORGANISM

Brassica oleracea
Brassica oleracea
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
rosids; eurosids II; Brassicales; Brassicaceae; Brassica;

REFERENCE

AUTHORS

Ayele,M., Haas,B.J., Kumar,N., Wu,H., Xiao,Y., Van Aken,S.,
Utterback,T.R., Wortman,J.R., White,O.R. and Town,C.D.
Whole genome shotgun sequencing of Brassica oleracea and its
application to gene discovery and annotation in Arabidopsis
Genome Res. 15 (4), 487-495 (2005)
15805490

TITLE

JOURNAL

PUBMED

COMMENT

Other GSSs: BOND17TF

Contact: Chris Town

TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA.

Tel: 301-838-3523

Fax: 301-838-0208

Email: cdtown@tigr.org

DNA is from a doubled haploid provided by Tom Osborn.

Seq primer: TR

Class: Sheared ends.

Location/Qualifiers

1..918

/organism="Brassica oleracea"

/mol_type="genomic DNA"

/strain="TO1000DH3"

/db_xref="taxon:3712"

/clone="BOND17"

/clone_lib="BO.1.6.2_KB tot"

/note=Vector: pHS1; Site 1: BstXI; 1.6-2 kb sheared

total DNA inserted into pHS1 using BstXI linkers"

FEATURES

source

Query Match

Best Local Similarity

Matches

918; Conservative

0; Mismatches

0; Indels

0; Gaps

0;

ORIGIN

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

| RESULT 4 | CNS06X6V | 988 bp | DNA | linear | GSS 06-JUL-2001 |
|-------------|--|---|--------|---------------|-------------------|
| LOCUS | T7 end of clone AX0AA038H01 of library AX0AA from strain CBS 7064 | | | | |
| DEFINITION | of <i>Pichia farinosa</i> , genomic survey sequence. | | | | |
| ACCESSION | AL419357 | | | | |
| VERSION | AL419357.1 | GI:12020535 | | | |
| KEYWORDS | GSS. | | | | |
| SOURCE | <i>Pichia farinosa</i> | | | | |
| ORGANISM | <i>Pichia farinosa</i> | | | | |
| REFERENCE | Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; <i>Pichia</i> . | | | | |
| AUTHORS | 1 (bases 1 to 988) | | | | |
| | Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G., Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S., de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B., Malpertuy,A., Neuveglise,C., Osier-Kalogeropoulos,O., Potier,S., Saurin,W., Tekai,F., Toffano-Nioche,C., Wesolowski-Louvel,M., Wincker,P. and Weissenbach,J. | | | | |
| TITLE | Genomic exploration of the hemiascomycetous yeasts: 1. A set of yeast species for molecular evolution studies | | | | |
| JOURNAL | FEBS Lett. 487 (1), 3-12 (2000) | | | | |
| PUBMED | 11152876 | | | | |
| REFERENCE | 2 (bases 1 to 988) | | | | |
| AUTHORS | de Montigny,J., Spehner,C., Souciet,J., Tekai,F., Dujon,B., Wincker,P., Artiguenave,F. and Potier,S. | | | | |
| TITLE | Genomic exploration of the hemiascomycetous yeasts: 15. <i>Pichia sorbitophila</i> | | | | |
| JOURNAL | FEBS Lett. 487 (1), 87-90 (2000) | | | | |
| PUBMED | 11152890 | | | | |
| REFERENCE | 3 (bases 1 to 988) | | | | |
| AUTHORS | Genoscope. | | | | |
| TITLE | Direct Submission | | | | |
| JOURNAL | Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage, 2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr) | | | | |
| COMMENT | This GSS is part of a random genomic sequencing program of thirteen yeast species: <i>Saccharomyces bayanus</i> var. <i>uvarum</i> , <i>Saccharomyces exiguus</i> , <i>Saccharomyces servazzii</i> , <i>Zygosaccharomyces rouxii</i> , <i>Saccharomyces kluyveri</i> , <i>Kluyveromyces thermotolerans</i> , <i>Kluyveromyces lactis</i> var. <i>lactis</i> , <i>Kluyveromyces marxianus</i> var. <i>marxianus</i> , <i>Pichia angusta</i> , <i>Debaryomyces hansenii</i> var. <i>hansenii</i> , <i>Pichia sorbitophila</i> , <i>Candida tropicalis</i> and <i>Yarrowia lipolytica</i> . Genomic inserts of 3 to 5 kb were prepared and both extremities were sequenced. See keywords for description of this sequence and for the sequence of the other extremity of this insert. | | | | |
| FEATURES | Location/Qualifiers | | | | |
| source | 1..988 | | | | |
| | /organism="Pichia farinosa" | | | | |
| | /mol_type="genomic DNA" | | | | |
| | /strain="CBS 7064" | | | | |
| | /db_xref="taxon:4920" | | | | |
| | /clone="AX0AA038H01" | | | | |
| | /clone_lib="AX0AA" | | | | |
| | /note="end : T7" | | | | |
| ORIGIN | | | | | |
| Query Match | 14.9%; | Score 719.2; | DB 11; | Length 988; | |
| Best Local | Similarity 99.0%; | Pred. No. 7.3e-190; | | | |
| Matches | 721; | Conservative | 2; | Mismatches 5; | Indels 0; Gaps 0; |
| Qy | 3636 | CAATTCGATTAGAAAACTCATCGAGCATCAATGAACTGCAATTATTCATATCAGG | 3695 | | |
| Db | 261 | CAGTACTGATTAGAAAAACTCATCGAGCATCAATGAACTGCAATTATTCATATCAGG | 320 | | |
| Qy | 3696 | ATTATCAATACCATTTTGGAAAAAGCCGTTTCTGTAATGAAGGAGAAACTCACCGAG | 3755 | | |
| Db | 321 | ATTATCAATACCTTTATTTTGGAAAAAGCCGTTTCTGTAATGAAGGAGAAACTCACCGAG | 380 | | |
| Qy | 3756 | GCAGTTCATAGGATGCGAAGATCCTCGGTATCGGTCTGCGATTCGCAATTCGCTCCCAATC | 3815 | | |
| Db | 381 | GCAGTTCATAGGATGCGAAGATCCTCGGTATCGGTCTGCGATTCGCAATTCGCTCCCAATC | 440 | | |

```
High quality sequence start: 16
High quality sequence stop: 627.
Location/Qualifiers
1..889
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:6436303"
/lab_host="DH10B"
/clone_lib="NIH_MGC_137"
Note="Organ: pancreas; Vector: pSPORT1; Site 1: SalI;
Site 2: NotI; Library consists of a pool of clones
rearrayed from the following libraries: Melton normalized
mixed mouse pancreas 1 NI-MMS1, Amplified Melton mouse
islets 1 Misi-A, and Kaestner mgn3 wt. Clones rearrayed in
the laboratory of K. Kaestner (University of
Pennsylvania). Note: this is a NIH_MGC Library."
ORIGIN
Query Match 14.6%; Score 704.8; DB 6; Length 889;
Best Local Similarity 97.4%; Pred. No. 7.8e-186;
Matches 738; Conservative 0; Mismatches 17; Indels 3; Gaps 2;
QY 3582 GCGCGCGTCCGCTCAAGTCAGGTAAATGCTCTGCCAGTGTACAAACCAATTAACCAATTC 3641
DB 120 GCGCGCGTCCGCTCAAGTCAGGTAAATGCTCTGCCAGTGTACAAACCAATTAACCAATTC 179
QY 3642 TGATTAGAAAACTCATCGAGCATCAAAATGAAACTGCAATTTATTCATATCATGAGTATTC 3701
DB 180 TGATTAGAAAACTCATCGAGCATCAAAATGAAACTGCAATTTATTCATATCATGAGTATTC 239
QY 3702 AATACCATATTTTGAAGAAGCGTTTCTGTAATGAAGAGAGAAACTCAACGAGGAGTT 3761
DB 240 AATACCATATTTTGAAGAAGCGTTTCTGTAATGAAGAGAGAAACTCAACGAGGAGTT 299
QY 3762 CCATAGGATGCAAGATCTGTATCGGTCTCGGATTCGACTCGTCCAAATCAATACATACA 3821
DB 300 CCATAGGATGCAAGATCTGTATCGGTCTCGGATTCGACTCGTCCAAATCAATACATACA 359
QY 3822 ACCTATTAAATTTCCCTCGTCAAAATGAAGTTTCAAGTGAAGAAATCAACATGAGTAC 3881
DB 360 ACCTATTAAATTTCCCTCGTCAAAATGAAGTTTCAAGTGAAGAAATCAACATGAGTAC 419
QY 3882 GACTGAATCCGCTGAGATGCAAAAGCTTATGCAATTTCTTCCAGACTTGTTCACAGG 3941
DB 420 GACTGAATCCGCTGAGATGCAAAAGCTTATGCAATTTCTTCCAGACTTGTTCACAGG 479
QY 3942 CCAGCCATTACGCTCATCAAAATCACTCGCATCAACCAACCGTTATTCATTCGTGA 4001
DB 480 CCAACCATTAACGCTCATCAAAATCACTCGCATCAACCAACCGTTATTCATTCGTGA 539
QY 4002 TTGCGCTGAGCGAGACGAAATACGCGATCGTGTAAAGAGCAATTAACAAACAGGAAT 4061
DB 540 TTGCGCTGAGCGAGACGAAATACGCGATCGTGTAAAGAGCAATTAACAAACAGGAAT 599
QY 4062 CGAATGCAACCGGCGAGCAACATCTGCGAGGATCAACAAATTTTCACTGATACAGG 4121
DB 600 CGAATGCAACCGGCGAGCAACATCTGCGAGGATCAACAAATTTTCACTGATACAGG 659
QY 4122 ATATTCTTCTAATACCTGGAATCTGTTTTTCCGGGATCGAGTGGTGAATACCATGC 4181
DB 660 ATATTCTTCTAATACCTGGAATCTGTTTTTCCGGGATCGAGTGGTGAATACCATGC 719
QY 4182 ATCATCAGGATGACGATTAATGCTGATGTCGGAAGAGCATATAAATTCGTCAGCA 4241
DB 720 ATCATCAGGATGACGATTAATGCTGATGTCGGAAGAGCATATAAATTCGTCAGCA 779
QY 4242 GTTTAGTCTGACCAATCTCATCTGTAACATCATTTGGCAACCGCTTACCTTTGCCATG-TTTC 4300
DB 780 GTTTAGTCTGACCAATCTCATCTGTAACATCATTTGGCAACCGCTTACCTTTGGCATGTTTCA 839
QY 4301 GAAACAATCTGG--CGCATCGGGCTTCCCATACATC 4336
|||||
```

```
Db 840 AAAACAACCTCTGGCGCATCCGGGCTTCCCTTACAATC 877
RESULT 6
BH600610/c
LOCUS
DEFINITION
BH600610 675 bp DNA linear GSS 15-DB-2001
BOHKQ87TF BOHK Brassica oleracea genomic clone BOHKQ87, genomic
survey sequence.
ACCESSION
BH600610 GI:17853056
VERSION
BH600610
KEYWORDS
GSS.
SOURCE
Brassica oleracea
Brassica oleracea
ORGANISM
Brassica oleracea
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE
1 (bases 1 to 675)
Avele,M., Haas,B.J., Kumar,N., Wu,H., Xiao,Y., Van Aken,S.,
Utterback,T.R., Wortman,J.R., White,O.R. and Town,C.D.
Whole genome shotgun sequencing of Brassica oleracea and its
application to gene discovery and annotation in Arabidopsis
Genome Res. 15 (4), 487-495 (2005)
15805490
CONTACT: Chris Town
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: 1F
Class: sheared ends.
Location/Qualifiers
1..675
/organism="Brassica oleracea"
/mol_type="genomic DNA"
/strain="T01000DH3"
/db_xref="taxon:3712"
/clone="BOHKQ87"
/clone_lib="BOHK"
/notes="Vector: pHS1; Site 1: BstXI; 2-3 kb sheared
genomic DNA inserted into pHS1 using BstXI linkers"
FEATURES
source
ORIGIN
Query Match 13.9%; Score 670.2; DB 9; Length 675;
Best Local Similarity 99.6%; Pred. No. 3.8e-176;
Matches 672; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 2047 TCGTCCGCGCACTTATGACTGTCTTTTATCATGCAACTCGTAGGACAGGTGCGGCGAG 2106
DB 675 TCGTCCGCGCACTTATGACTGTCTTTTATCATGCAACTCGTAGGACAGGTGCGGCGAG 616
QY 2107 CGCTCTGGGTCAATTTTCGGCGAGGACCGCTTTTCGCTGGAGCGGACGATGATCGGCTGT 2166
DB 615 CGCTCTGGGTCAATTTTCGGCGAGGACCGCTTTTCGCTGGAGCGGACGATGATCGGCTGT 556
QY 2167 CGCTTGGGTATTCGGGAATCTTGACGCGCTTCGCTCAAGCCCTTCGCTCACTGCTCCGCCA 2226
DB 555 CGCTTGGGTATTCGGGAATCTTGACGCGCTTCGCTCAAGCCCTTCGCTCACTGCTCCGCCA 496
QY 2227 CCAACGTTTCGGCGAGAAAGCAGGCATATTATCGCGGATCGGCGGCGGCGGCTGGGCT 2286
DB 495 CCAACGTTTCGGCGAGAAAGCAGGCATATTATCGCGGCGGCGGCGGCTGGGCT 436
QY 2287 ACGTCTTGTCTGGCTTTCGGAGCGGAGCTGGATGGCCCTCCCATTTATGATTTCTCTCG 2346
DB 435 ACGTCTTGTCTGGCTTTCGGAGCGGAGCTGGATGGCCCTCCCATTTATGATTTCTCTCG 376
QY 2347 CTTCCGGCGGATCGGGATGCGCGTTCGAGGCCATCTGTCAGGCGAGGTAGATGACG 2406
DB 375 CTTCCGGCGGATCGGGATGCGCGTTCGAGGCCATCTGTCAGGCGAGGTAGATGACG 316
QY 2407 ACCATCAGGACAGCTTCAAGGATCGCTCGCGGCTCTTTACCAGCGCTAACTTCGATCATTTG 2466
|||||
```

Db 315 ACCATCAGGACAGCTTCAAGGATCGCTCGGGCTCTTACCAGCCTCACTTCGATCATTTG 256
QY 2467 GACCGCTGATCGTCAAGCGGATTTATGCGGCTTCGGCGAGACATGGAACGGGTTGGCAT 2526
Db 255 GACCGCTGATCGTCAAGCGGATTTATGCGGCTTCGGCGAGACATGGAACGGGTTGGCAT 196
QY 2527 GGATGTAGGCGCGCCCTATACCTTGTCTGCTCCCGCGTTGCGTTCGGGTGATGGA 2586
Db 195 GGATGTAGGCGCGCCCTATACCTTGTCTGCTCCCGCGTTGCGTTCGGGTGATGGA 136
QY 2587 GCGCGCCACCTCGACCTGAATGGAAGCGCGGACCTCGCTAACGGANTTCAACACTCC 2646
Db 135 GCGCGCCACCTCGACCTGAATGGAAGCGCGGACCTCGCTAACGGANTTCAACACTCC 76
QY 2647 AAGAAATGGAGCAATCAATTTCTCGGAGAACTGTGAATGGCAAAACAACCTTTGGCA 2706
Db 75 AAGAAATGGAGCAATCAATTTCTCGGAGAACTGTGAATGGCAAAACCGGCAATTTGCA 16
QY 2707 GAACATATCCATCGC 2721
Db 15 GAACATATCCATCGC 1

RESULT 7
CG411076 663 bp DNA linear GSS 01-NOV-2003
LOCUS RM391 Lxx transposon mutant library Leifsonia xyli subsp. xyli
DEFINITION genomic, genomic survey sequence.
ACCESSION CG411076
VERSION CG411076.1 GI:38142787
KEYWORDS GSS.
SOURCE Leifsonia xyli subsp. xyli
ORGANISM Leifsonia xyli subsp. xyli
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Micrococciaceae; Microbacteriaceae; Leifsonia.
REFERENCE 1 (bases 1 to 663)
AUTHORS Brumbley, S.M., Petrasovits, L.A., Murphy, R.M., Nagel, R.J.,
Candy, J.M. and Hermann, S.R.
TITLE Establishment of a functional genomics platform for Leifsonia xyli
subsp. xyli
JOURNAL Mol. Plant Microbe Interact. 17 (2), 175-183 (2004)
PUBMED 14964531
COMMENT Contact: Brumbley SM
DNPNC
BSES Ltd
50 Meiers Rd, Indooroopilly, Brisbane, 4068, Australia
Tel: 61 7 3331 3333
Fax: 61 7 3871 0383
Email: sbrumbley@bSES.org.au
Class: transposon insertion site.
Location/Qualifiers
1. 663
/organism="Leifsonia xyli subsp. xyli"
/mol_type="genomic DNA"
/sub_species="xyli"
/db_xref="taxon:59736"
/clone_lib="Lxx transposon mutant library"
/note="Vector: pUCD623; Lxx transposon mutants were
generated by electroporating pUCD623 containing the
transposon Tn4431 into Lxx cells grown from 10-15 days in
S8 broth containing 0.1% glycine."

FEATURES
source

Query Match 13.7%; Score 658; DB 10; Length 663;
Best Local Similarity 100.0%; Pred. No. 1e-172;
Matches 658; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2118 ATTTTCGGCGAGACCGCTTTTCGCTGGAGCGGACGATCGGCTGTCGTTGCGGTA 2177
Db 1 ATTTTCGGCGAGACCGCTTTTCGCTGGAGCGGACGATCGGCTGTCGTTGCGGTA 60

ORIGIN

Query Match 13.7%; Score 658; DB 10; Length 663;
Best Local Similarity 100.0%; Pred. No. 1e-172;
Matches 658; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2118 ATTTTCGGCGAGACCGCTTTTCGCTGGAGCGGACGATCGGCTGTCGTTGCGGTA 2177
Db 1 ATTTTCGGCGAGACCGCTTTTCGCTGGAGCGGACGATCGGCTGTCGTTGCGGTA 60
QY 2178 TTCGGAAATTCGACGCCCTCGCTCAAGCCTTCGTCACCTGCTCCCGCCACCAACGTTTC 2237

Db 61 TTTCGAATTTTTCAGCGCCTCGCTCAAGCCTTCGTCACCTGCTCCCGCACCAACGTTTC 120
QY 2238 GGCAGAGAAGCAGGCATTTATCGCGGATGCGCGCGGCGGCGCTGGGCTAGCTTTCGTCG 2297
Db 121 GGCAGAGAAGCAGGCATTTATCGCGGATGCGCGCGGCGGCGCTGGGCTAGCTTTCGTCG 180
QY 2298 GCGTTTCGCGAGCGGAGGCTGGATGCGCTTCCCAATTAATGATTTCTTCGCTTCGGCGGCG 2357
Db 181 GCGTTTCGCGAGCGGAGGCTGGATGCGCTTCCCAATTAATGATTTCTTCGCTTCGGCGGCG 240
QY 2358 ATCGGGATGCGCGGCTTTCAGGCCCATGCTGTCCAGGCAAGGTAGATGACGACCATCAGGA 2417
Db 241 ATCGGGATGCGCGGCTTTCAGGCCCATGCTGTCCAGGCAAGGTAGATGACGACCATCAGGA 300
QY 2418 CAGCTTCAAGATCGCTCGGCGCTTTACAGCGCTTAATTCGATCATTTGGACCGCTGATC 2477
Db 301 CAGCTTCAAGATCGCTCGGCGCTTTACAGCGCTTAATTCGATCATTTGGACCGCTGATC 360
QY 2478 GTCACGGCGATTTATGCGCGCTCGCGGAGCACATGGAACGGGTTGGCATGATTCGTAGGC 2537
Db 361 GTCACGGCGATTTATGCGCGCTCGCGGAGCACATGGAACGGGTTGGCATGATTCGTAGGC 420
QY 2538 GCGCGCCTATACCTTGTCTGCTCCCGCGTTGCGGTGCGGTGCGATGGAGCGGCGCACCC 2597
Db 421 GCGCGCCTATACCTTGTCTGCTCCCGCGTTGCGGTGCGGTGCGATGGAGCGGCGCACCC 480
QY 2598 TCGACCTGAATGGAAGCGCGGCGGCACTTCGCTAACGGAATTCACCTCCAGAAATTTGAG 2657
Db 481 TCGACCTGAATGGAAGCGCGGCGGCACTTCGCTAACGGAATTCACCTCCAGAAATTTGAG 540
QY 2658 CCAATCAATTTCTCGGAGAACTGTGAATGCGAAACCAACCTTGGCAGAACATATCCA 2717
Db 541 CCAATCAATTTCTCGGAGAACTGTGAATGCGAAACCAACCTTGGCAGAACATATCCA 600
QY 2718 TCGGCTCGCGCATCTTCAGAGCGCGGCGCATCTCGGGGAGCGCTTGGGTCCTG 2775
Db 601 TCGGCTCGCGCATCTTCAGAGCGCGGCGCATCTCGGGGAGCGCTTGGGTCCTG 658

RESULT 8

AJ762384 648 bp mRNA linear EST 03-JUN-2005
LOCUS AJ762384 G00006 Gerbera hybrid cv. 'Terra Regina' cDNA clone
DEFINITION G000060001D12F1, mRNA sequence.

ACCESSION AJ762384

VERSION AJ762384.1 GI:62943595

KEYWORDS EST.

SOURCE Gerbera hybrid cv. 'Terra Regina'

ORGANISM Gerbera hybrid cv. 'Terra Regina',

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
asterids; campanulids; Asterales; Asteraceae; Mutisioideae;
Mutisieae; Gerbera.

REFERENCE 1 (bases 1 to 648)

AUTHORS Laitinen, R.A., Imanen, J., Auvinen, P., Rudd, S., Alatalo, E.R.,

Paulin, L., Ainasoja, M., Kotilainen, M., Koskela, S., Teeri, T.H. and
Elomaa, P.

Analysis of the floral transcriptome uncovers new regulators of
organ determination and gene families related to flower organ
differentiation in Gerbera hybrida (Asteraceae)

Genome Res. 15 (4), 475-486 (2005)

JOURNAL 15781570

PUBMED Contact: Alatalo ER

COMMENT Institute of Biotechnology

Viikinkaari 9, P.O. Box 56, 00014 University of Helsinki, FINLAND.

FEATURES Location/Qualifiers

source 1. 648

/organism="Gerbera hybrid cv. 'Terra Regina'"

/mol_type="mRNA"

/cultivar="Terra Regina"

/db_xref="taxon:226891"

/clone="G000060001D12F1"

/tissue_type="flower stem"
/clone_lib="G00006"

ORIGIN

```

Query Match      13.5%; Score 648; DB 1; Length 648;
Best Local Similarity 100.0%; Pred. No. 6.5e-170;
Matches 648; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1334 TCATGTTTGACAGCTTATCATCGATAAGCTTTAAATGCGGTAGTTTATCACAGTTAAATG 1393
Db 1 TCATGTTTGACAGCTTATCATCGATAAGCTTTAAATGCGGTAGTTTATCACAGTTAAATG 60

QY 1394 CTAACGCGAGTCAGGACCGTGTATCAAAATCTAAACATGGCTCATCGTCTCATCTCGGCAC 1453
Db 61 CTAACGCGAGTCAGGACCGTGTATCAAAATCTAAACATGGCTCATCGTCTCATCTCGGCAC 120

QY 1454 CGTCACCCCTGGATGCTGTAGGCATAGGCTTGGTTATGCGGTACTGCGGGGCTCTTCGG 1513
Db 121 CGTCACCCCTGGATGCTGTAGGCATAGGCTTGGTTATGCGGTACTGCGGGGCTCTTCGG 180

QY 1514 GGATATGCTCAATTCCTATGCGCAGCATCGCCAGTCACTATGCGGTGCTGCTAGCGCTATATGC 1573
Db 181 GGATATGCTCAATTCCTATGCGCAGCATCGCCAGTCACTATGCGGTGCTGCTAGCGCTATATGC 240

QY 1574 GTTGATGCAATTTCTATGCGCAGCATCGCCAGTCACTATGCGGTGCTGCTAGCGCTATATGC 1633
Db 241 GTTGATGCAATTTCTATGCGCAGCATCGCCAGTCACTATGCGGTGCTGCTAGCGCTATATGC 300

QY 1634 CCCAGTCTGCTGCTTACGCGGACGATCGTGGCGGACATCGGATCGGATCGGACAC 1693
Db 301 CCCAGTCTGCTGCTTACGCGGACGATCGTGGCGGACATCGGATCGGATCGGACAC 360

QY 1694 ACCGCTCTGCTGCTTACGCGGACGATCGTGGCGGACATCGGATCGGATCGGACAC 1753
Db 361 ACCGCTCTGCTGCTTACGCGGACGATCGTGGCGGACATCGGATCGGATCGGACAC 420

QY 1754 TGGGCTGCTGGCGCTATATGCGGACATCACCGATGGGAAGATCGGCTCGCACTT 1813
Db 421 TGGGCTGCTGGCGCTATATGCGGACATCACCGATGGGAAGATCGGCTCGCACTT 480

QY 1814 CGGGCTATGAGCGCTTCTTTCGGCGTGGGTATGTTGCGGCGCCCGGCGGGGACT 1873
Db 481 CGGGCTATGAGCGCTTCTTTCGGCGTGGGTATGTTGCGGCGCCCGGCGGGGACT 540

QY 1874 GTTGGGCGCCATCTCTTGCATGACCATCTCTTGGCGGCGGTGCTCAACGGCCTCAA 1933
Db 541 GTTGGGCGCCATCTCTTGCATGACCATCTCTTGGCGGCGGTGCTCAACGGCCTCAA 600

QY 1934 CCTACTACTGGGCTGCTTCTTAATGCGAGATCGCATTAAGGAGAGCG 1981
Db 601 CCTACTACTGGGCTGCTTCTTAATGCGAGATCGCATTAAGGAGAGCG 648

```

RESULT 9

```

AJ762542
LOCUS
DEFINITION
G0000600013D05F1, mRNA linear EST 03-JUN-2005
AJ762542
VERSION
AJ762542.1 GI:62943753
KEYWORDS
SOURCE
ORGANISM
Gerbera hybrid cv. 'Terra Regina'
Gerbera hybrid cv. 'Terra Regina'
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
asterids; campanulids; Asterales; Asteraceae; Mutisioideae;
Mutisieae; Gerbera
1 (bases 1 to 648)
Laitinen,R.A., Immanen,J., Auvinen,P., Rudd,S., Alatalo,E.R.,
Paulin,L., Ainasoja,M., Kotilainen,M., Koskela,S., Teeri,T.H. and
Elomaa,P.
Analysis of the floral transcriptome uncovers new regulators of
organ determination and gene families related to flower organ

```

REFERENCE

```

AUTHORS
Laitinen,R.A., Immanen,J., Auvinen,P., Rudd,S., Alatalo,E.R.,
Paulin,L., Ainasoja,M., Kotilainen,M., Koskela,S., Teeri,T.H. and
Elomaa,P.
TITLE
Analysis of the floral transcriptome uncovers new regulators of
organ determination and gene families related to flower organ

```

RESULT 10

```

BF632491
LOCUS
DEFINITION
NF027F11DT1F1092 Drought Medicago truncatula cDNA clone NF027F11DT
S', mRNA sequence.
ACCESSION
BF632491
VERSION
BF632491.1 GI:11896649

```

differentiation in Gerbera hybrida (Asteraceae)
Genome Res. 15 (4), 475-486 (2005)

JOURNAL

PUBMED

COMMENT

Contact: Alatalo ER
Institute of Biotechnology
Viikinkaari 9, P.O. Box 56, 00014 University of Helsinki, FINLAND.
Location/Qualifiers
1..648

FEATURES

source

/organism="Gerbera hybrid cv. 'Terra Regina'"
/mol_type="mRNA"
/cultivar="Terra Regina"
/db_xref="taxon:226891"
/clone="G0000600013D05F1"
/tissue_type="flower stem"
/clone_lib="G00006"

ORIGIN

```

Query Match      13.5%; Score 648; DB 1; Length 648;
Best Local Similarity 100.0%; Pred. No. 6.5e-170;
Matches 648; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1334 TCATGTTTGACAGCTTATCATCGATAAGCTTTAAATGCGGTAGTTTATCACAGTTAAATG 1393
Db 1 TCATGTTTGACAGCTTATCATCGATAAGCTTTAAATGCGGTAGTTTATCACAGTTAAATG 60

QY 1394 CTAACGCGAGTCAGGACCGTGTATCAAAATCTAAACATGGCTCATCGTCTCATCTCGGCAC 1453
Db 61 CTAACGCGAGTCAGGACCGTGTATCAAAATCTAAACATGGCTCATCGTCTCATCTCGGCAC 120

QY 1454 CGTCACCCCTGGATGCTGTAGGCATAGGCTTGGTTATGCGGTACTGCGGGGCTCTTCGG 1513
Db 121 CGTCACCCCTGGATGCTGTAGGCATAGGCTTGGTTATGCGGTACTGCGGGGCTCTTCGG 180

QY 1514 GGATATGCTCAATTCCTATGCGCAGCATCGCCAGTCACTATGCGGTGCTGCTAGCGCTATATGC 1573
Db 181 GGATATGCTCAATTCCTATGCGCAGCATCGCCAGTCACTATGCGGTGCTGCTAGCGCTATATGC 240

QY 1574 GTTGATGCAATTTCTATGCGCAGCATCGCCAGTCACTATGCGGTGCTGCTAGCGCTATATGC 1633
Db 241 GTTGATGCAATTTCTATGCGCAGCATCGCCAGTCACTATGCGGTGCTGCTAGCGCTATATGC 300

QY 1634 CCCAGTCTGCTGCTTACGCGGACGATCGTGGCGGACATCGGATCGGATCGGACAC 1693
Db 301 CCCAGTCTGCTGCTTACGCGGACGATCGTGGCGGACATCGGATCGGATCGGACAC 360

QY 1694 ACCGCTCTGCTGCTTACGCGGACGATCGTGGCGGACATCGGATCGGATCGGACAC 1753
Db 361 ACCGCTCTGCTGCTTACGCGGACGATCGTGGCGGACATCGGATCGGATCGGACAC 420

QY 1754 TGGGCTGCTGGCGCTATATGCGGACATCACCGATGGGAAGATCGGCTCGCACTT 1813
Db 421 TGGGCTGCTGGCGCTATATGCGGACATCACCGATGGGAAGATCGGCTCGCACTT 480

QY 1814 CGGGCTATGAGCGCTTCTTTCGGCGTGGGTATGTTGCGGCGCCCGGCGGGGACT 1873
Db 481 CGGGCTATGAGCGCTTCTTTCGGCGTGGGTATGTTGCGGCGCCCGGCGGGGACT 540

QY 1874 GTTGGGCGCCATCTCTTGCATGACCATCTCTTGGCGGCGGTGCTCAACGGCCTCAA 1933
Db 541 GTTGGGCGCCATCTCTTGCATGACCATCTCTTGGCGGCGGTGCTCAACGGCCTCAA 600

QY 1934 CCTACTACTGGGCTGCTTCTTAATGCGAGATCGCATTAAGGAGAGCG 1981
Db 601 CCTACTACTGGGCTGCTTCTTAATGCGAGATCGCATTAAGGAGAGCG 648

```

KEYWORDS EST.
SOURCE Medicago truncatula (barrel medic)
ORGANISM Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
Medicago.
1 (bases 1 to 659)
Torres-Jerez, I., Scott, A.D., Harris, A.R., Gonzales, R.A., Bell, C.J.,
Flores, H.R., Inman, J.T., Weller, J.W. and May, G.D.
Expressed Sequence Tags from the Samuel Roberts Noble Foundation
Medicago truncatula drought library
Unpublished (2000)
Contact: May GD
Plant Biology Division
The Samuel Roberts Noble Foundation
2510 Sam Noble Parkway, Ardmore, OK 73402, USA
Tel: 580 224 6650
Fax: 580 224 6692
Email: gdmay@noble.org
Insert Length: 659 Std Error: 0.00
Plate: 027 row: F column: 11
Seq primer: TCACACAGGAACAGCTATGAC.
FEATURES
Location/Qualifiers
1..659
/organism="Medicago truncatula"
/mol_type="mRNA"
/db_xref="taxon:3880"
/clone="NF027F11DT"
/tissue_type="Plantlets"
/dev_stage="Pooled timepoints"
/clone_lib="Drought"
/note="Vector: Lambda Zap; Contains a mixture of entire
plantlets harvested in a series of days-post-watering
timepoints. "
ORIGIN
Query Match 13.4%; Score 646; DB 2; Length 659;
Best Local Similarity 99.8%; Pred. No. 2.4e-169;
Matches 657; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
QY 1329 AATTCCTCATGTTTGACAGCTTATCATGATAGCTTTAATCGGTAGTTTATCACAGTTA 1388
DB 1 AATTCCTCATGTTTGACAGCTTATCATGATAGCTTTAATCGGTAGTTTATCACAGTTA 60
QY 1389 AATTCCTAAGCGAGTCAGGACCGGTGTATGAATCTAAATGCGCTCATGTCATCTC 1448
DB 61 AATTCCTAAGCGAGTCAGGACCGGTGTATGAATCTAAATGCGCTCATGTCATCTC 120
QY 1449 GGCACCGTCACCGTGGATGCTTAGGCATAGGCTTGGTTATGCCGGTACTGCCGGGCTC 1508
DB 121 GGCACCGTCACCGTGGATGCTTAGGCATAGGCTTGGTTATGCCGGTACTGCCGGGCTC 180
QY 1509 TTGCGGGATATCGTCCATTCGACAGCATCGCAGTCATATGCGGTGCTGCTAGCGCTA 1568
DB 181 TTGCGGGATATCGTCCATTCGACAGCATCGCAGTCATATGCGGTGCTGCTAGCGCTA 240
QY 1569 TATGCGGTTGATCAATTTCTATGCGCACCGGTTCTCGGAGCACTGTCGACCGCTTTCGC 1628
DB 241 TATGCGGTTGATCAATTTCTATGCGCACCGGTTCTCGGAGCACTGTCGACCGCTTTCGC 300
QY 1629 CCGCCGCCAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1688
DB 301 CCGCCGCCAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 360
QY 1689 ACCACACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1748
DB 361 ACCACACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 420
QY 1749 ACA-GGTGCGGTGCTGCGGCTTATATCGCCGACATCAACCGATGGGGAGATCGGGCTCG 1807
DB 421 ACAGGGTGGGTGCTGCGGCTTATATCGCCGACATCAACCGATGGGGAGATCGGGCTCG 480

QY 1808 CCACCTCGGGCTCATGAGCGCTTGTTCGCGGTGGGTATGTTGGCAGGCCCGCGGCGG 1867
DB 481 CCACCTCGGGCTCATGAGCGCTTGTTCGCGGTGGGTATGTTGGCAGGCCCGCGGCGG 540
QY 1868 GGGACTGTTGGCGGCCATCTCTTTCGATGACCAATTCCTTTCGCGGCGGGTGTCTCAACGG 1927
DB 541 GGGACTGTTGGCGGCCATCTCTTTCGATGACCAATTCCTTTCGCGGCGGGTGTCTCAACGG 600
QY 1928 CCTCAACTACTACTGCGGCTCTCTTAATGCAGGAGTCGATGAAGGAGAGCGCTCGA 1985
DB 601 CCTCAACTACTACTGCGGCTCTCTTAATGCAGGAGTCGATGAAGGAGAGCGCTCGA 658
RESULT 11
BQ752029 645 bp mRNA linear EST 18-JUL-2002
LOCUS EST632592 DSCT Colletotrichum trifolii cDNA clone pDSCT9-44, mRNA
DEFINITION sequence.
ACCESSION BQ752029
VERSION BQ752029.1 GI:21907434
KEYWORDS EST:
SOURCE Colletotrichum trifolii
ORGANISM Colletotrichum trifolii
REFERENCE 1 (bases 1 to 645)
AUTHORS Samac, D.A., Dickman, M., Town, C.D., Van Aken, S., Utterback, T.,
Cheung, F. and Fraser, C.M.
ESTs from mycelia of Colletotrichum trifolii race 1
TITLE Unpublished (2002)
JOURNAL Contact: Deborah A. Samac
COMMENT Department of Plant Pathology
University of Minnesota
495 Borlaug Hall, 1991 Upper Buford Circle, St. Paul, MN 55108, USA
Tel: 612 625 1243
Fax: 612 649 5058
Email: debby@puccini.crl.umn.edu
TIGR sequence name: MTSAL44TV More information is available at:
www.medicago.org
Seq primer: (gtA AtA CgA CtC Act AtA ggg C).
FEATURES
Location/Qualifiers
1..645
/organism="Colletotrichum trifolii"
/mol_type="mRNA"
/strain="race 1"
/db_xref="taxon:5466"
/clone="pDSCT9-44"
/tissue_type="mycelia"
/dev_stage="Young, actively growing mycelia (3 days after
inoculation) grown in liquid culture (cutin minimal medium
containing 2% glucose)."
/lab_host="DH5alpha"
/clone_lib="DSCT"
/note="Vector: pBluescript SK+; Site 1: EcoRI; Site 2:
EcoRI; isolate: 2sp2; cDNA was ligated into Lambda gtl1 from
Stratagene and packaged using Gigapack packaging extracts.
An aliquot of the amplified library was used to transduce
E. coli Y1090 and phage DNA was purified from a liquid
lysate. The cDNA inserts were gel purified after EcoRI
digestion and ligated into pBluescript SK+. Aliquots of
the ligation were used to transform E. coli DH5alpha which
were plated onto medium with X-gal for selection of
recombinants."
ORIGIN
Query Match 13.4%; Score 645; DB 5; Length 645;
Best Local Similarity 100.0%; Pred. No. 4.5e-169;
Matches 645; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1334 TCATGTTTGACAGCTTATCATCGATAAGCTTTAATGCGGTAGTTTATCACAGTTAAATG 1393

```
Db 1 TCATGTTTGACAGCTTATCATCGATAGCTTTAAATGCGGTAGTTTATACAGATTAAATTTG 60
Qy 1394 CTAACGCGAGTCAGGACCGTGTATGAATCTAACAAATGCGGTATCGTATCTCTCGGCAC 1453
Db 61 CTAACGCGAGTCAGGACCGTGTATGAATCTAACAAATGCGGTATCGTATCTCTCGGCAC 120
Qy 1454 CGTACCCCTGGATGCTGTAGGCATAGGCTTGGTTATGCGGTATGCGGTATGCGGCGCTTTGGG 1513
Db 121 CGTACCCCTGGATGCTGTAGGCATAGGCTTGGTTATGCGGTATGCGGTATGCGGCGCTTTGGG 180
Qy 1514 GGATATCGTCCATTCGCGACAGCATCGCCAGTCACTATGCGGTGCTGTAGCGCTATATGTC 1573
Db 181 GGATATCGTCCATTCGCGACAGCATCGCCAGTCACTATGCGGTGCTGTAGCGCTATATGTC 240
Qy 1574 GTTGATGCAATTTCTATGGGACCCGCTTCTGGAGCATGTGCCAGCCGCTTTGGCGCGCG 1633
Db 241 GTTGATGCAATTTCTATGGGACCCGCTTCTGGAGCATGTGCCAGCCGCTTTGGCGCGCG 300
Qy 1634 CCCAGTCTGCTGCTGCTGCTACTTTGGAGCCACTATCGACTACGGATCATGGCGACAC 1693
Db 301 CCCAGTCTGCTGCTGCTGCTACTTTGGAGCCACTATCGACTACGGATCATGGCGACAC 360
Qy 1694 ACCGCTCTGTGGATCTCTAGCGCGAGCATGCTGTGCGCGCATCACCGCGGCCACAGG 1753
Db 361 ACCGCTCTGTGGATCTCTAGCGCGAGCATGCTGTGCGCGCATCACCGCGGCCACAGG 420
Qy 1754 TGGCGTGTGCGGCTTATATGCGCGACATCACCGATGGGAAAGATCGGGCTCGCCACTT 1813
Db 421 TGGCGTGTGCGGCTTATATGCGCGACATCACCGATGGGAAAGATCGGGCTCGCCACTT 480
Qy 1814 CGGCTCATAGAGCGCTTGTTCGGCGTGGGTATGCTGCGAGCCCGCTTGGCGCGGGAAT 1873
Db 481 CGGCTCATAGAGCGCTTGTTCGGCGTGGGTATGCTGCGAGCCCGCTTGGCGCGGGAAT 540
Qy 1874 GTTGGCGCCCATCTCTTGCATGACCATCTCTTGGCGCGGCTGCTCAACGGCTCAA 1933
Db 541 GTTGGCGCCCATCTCTTGCATGACCATCTCTTGGCGCGGCTGCTCAACGGCTCAA 600
Qy 1934 CCTACTGCGGCTGCTTCTTAATGACGAGGATGCGATAAGGGAGA 1978
Db 601 CCTACTGCGGCTGCTTCTTAATGACGAGGATGCGATAAGGGAGA 645
```

```
RESULT 12
CG804998/c
LOCUS
DEFINITION 118056B12.1BL_xl 1118 - RescueMu Grid S Zea mays genomic, genomic
survey sequence.
ACCESSION CG804998
VERSION CG804998.1 GI:38240986
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 681)
Walbot,V.
Maize genomic sequences found using engineered RescueMu transposon
Unpublished (2001)
Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Very probable ligation site of ends cut by single endonuclease.
Reverse complemented post-ligation sequence from source sequence.
Plate: 118056 row: 40
Class: transposon-tagged.
Location/Qualifiers
1 .681
FEATURES
source
```

```
/organism="Zea mays"
/mol_type="genomic DNA"
/cultivar="mixed background W23/A188/B73"
/db_xref="taxon:4577"
/tissue_type="leaf"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="1118 - RescueMu Grid S"
/note="Organ: leaf; Vector: RescueMu (engineered from
pBluescript backbone); Site 1: BamHI; Site 2: BglII;
RescueMu is a 4.9 kb, modified maize Mu transposon
designed to allow plasmid rescue from total genomic DNA.
Mu elements insert preferentially into transcription
units. For more information on RescueMu, go to the web
site 'www.zmdb.iaestate.edu' and follow the links for
'RescueMu.' Grid S was grown at San Diego in 2002. DNA was
extracted from leaf strips, double digested using BamHI
and BglII, and ligated to form circular plasmids. DH10B
cells were transformed and then screened on LB plates with
ampicillin."
```

ORIGIN

| Query Match | Best Local Similarity | Score | DB | Length | | | | | |
|-------------|-----------------------|--|------|------------|---|--------|---|------|---|
| Matches | 668 | Conservative | 0 | Mismatches | 5 | Indels | 2 | Gaps | 2 |
| Qy | 1706 | GATCCTCTACCGCGACGATCGTGGCGCGCATCACCGCGCGCACAGGTGCGGTGCTCGG | 1765 | | | | | | |
| Db | 681 | GATCCTCTACCGCGAGCGATCGTGGCGCGCATCACCGCGCGCACAGGTGCGGTGCTCGG | 622 | | | | | | |
| Qy | 1766 | CGCTTATATCGCGACATCAACGATGGGAAAGATCGGGCTCGCCACTTCGGGCTCATGAG | 1825 | | | | | | |
| Db | 621 | CGCTTATATCGCGACATCAACGATGGGAAAGATCGGGCTCGCCACTTCGGGCTCATGAG | 562 | | | | | | |
| Qy | 1826 | CGCTGTGTTTCGGCGTGGGTATGTTGGCAGGCCCGTGGCGGGGACACTGTTGGGCGCCAT | 1885 | | | | | | |
| Db | 561 | CGCTGTGTTTCGGCGTGGGTATGTTGGCAGGCCCGTGGCGGGGACACTGTTGGGCGCCAT | 502 | | | | | | |
| Qy | 1886 | CTCCTTGATGACACCATTCCTTTGGCGCGCGGTGCTCAACGGCTTCAACCTACTACTGGG | 1945 | | | | | | |
| Db | 501 | CTCCTTGATGACACCATTCCTTTGGCGCGCGGTGCTCAACGGCTTCAACCTACTACTGGG | 442 | | | | | | |
| Qy | 1946 | CTGCTTCTTAATGACGAGTGCATAGGGAGAGCGTGCAGCGATGCCCTTTGAGAGCCTT | 2005 | | | | | | |
| Db | 441 | CTGCTTCTTAATGACGAGTGCATAGGGAGAGCGTGCAGCGATGCCCTTTGAGAGCCTT | 382 | | | | | | |
| Qy | 2006 | CAACCCAGTACGCTCTTCGGTGGCGCGGGGATGACTATCGTCGCGCATTTATGAC | 2065 | | | | | | |
| Db | 381 | CAACCCAGTACGCTCTTCGGTGGCGCGGGGATGACTATCGTCGCGCATTTATGAC | 322 | | | | | | |
| Qy | 2066 | TGCTCTTTTATCATGCAACTCGTAGGACAGTGCAGGACAGCTCTCGGTCATTTTCGG | 2125 | | | | | | |
| Db | 321 | TGCTCTTTTATCATGCAACTCGTAGGACAGTGCAGGACAGCTCTCGGTCATTTTCGG | 262 | | | | | | |
| Qy | 2126 | CGAGGACCGCTTTCGCTGGAGCGGACGATGATCGGCTGTCGTTGCGGTATTCGGAAT | 2185 | | | | | | |
| Db | 261 | CGAGGACCGCTTTCGCTGGAGCGGACGATGATCGGCTGTCGTTGCGGTATTCGGAAT | 202 | | | | | | |
| Qy | 2186 | CTTGACGCGCTTCGCTCAAGCTTCGTCATCTGTTCCCGCCACCAAAAGTTTCGGCGAGAA | 2245 | | | | | | |
| Db | 201 | CTTGACGCGCTTCGCTCAAGCTTCGTCATCTGTTCCCGCCACCAAAAGTTTCGGCGAGTA | 142 | | | | | | |
| Qy | 2246 | GCAGGCCATTTATCGCGCATGGCGCGGATGGCGGCGGATGGCGGCTGGGCTACATCTTGGCGGTTCGC | 2305 | | | | | | |
| Db | 141 | GCAGGCCATTTATCGCGCATGGCGCGGATGGCGGCGGATGGCGGCTGGGCTACATCTTGGCGGTTCGC | 82 | | | | | | |
| Qy | 2306 | GACGCGAGGCTGGATGGCTTCCCATTTATGATTTCTTCTCGCTTCGCGCGGATCGGAT | 2365 | | | | | | |
| Db | 81 | GACGCGAGGCTGGATGGCTTCCCATTTATGATTTCTTCTCGCTTCGCGCGGATCGGAT | 24 | | | | | | |
| Qy | 2366 | GCCGCGTTGACGCG 2380 | | | | | | | |
| Db | 23 | GCCGCGTTGACGCG 9 | | | | | | | |


```

RESULT 13
BJ684207      675 bp      mRNA      linear      EST 25-JAN-2005
LOCUS        BJ684207 HCEST library Haplochromis chilotes cDNA clone no90g11,
DEFINITION   mRNA sequence.
ACCESSION    BJ684207
VERSION      BJ684207.1 GI:46527328
KEYWORDS     EST.
SOURCE       Haplochromis chilotes
ORGANISM     Haplochromis chilotes
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
              Acanthomorpha; Acanthopterygii; Percomorpha; Perciformes;
              Labroidae; Cichlidae; African cichlids; Pseudocrenilabrinae;
              Haplochromini; Haplochromis.
REFERENCE    1 (bases 1 to 675)
AUTHORS      Watanabe M., Kobayashi N., Shin-i, T., Horiike, T., Tateno, Y.,
              Kohara, Y. and Okada, N.
TITLE        Extensive analysis of ORF sequences from two different cichlid
              species in Lake Victoria provides molecular evidence for a recent
              radiation event of the Victoria species flock: identity of EST
              sequences between Haplochromis chilotes and Haplochromis sp.
JOURNAL      Gene 343 (2), 263-269 (2004)
PUBMED       15588581
COMMENT      Contact: Tadasu Shin-i
              Center For Genetic Resource Information
              National Institute of Genetics
              1111 Yata, Mishima, Shizuoka 411-8540, Japan
              Tel: 81-559-81-6856
              Fax: 81-559-81-6855
              Email: tshini@genes.nig.ac.jp.
FEATURES     source
              1..675
              /organism="Haplochromis chilotes"
              /mol_type="mRNA"
              /db_xref="taxon:257977"
              /clone="no90g11"
              /tissue_type="jaw"
              /dev_stage="varied"
              /clone_lib="HCEST library"
ORIGIN
Query Match      13.3%; Score 641; DB 3; Length 675;
Best Local Similarity 98.4%; Pred. No. 6.1e-168;
Matches 658; Conservative 0; Mismatches 10; Indels 1; Gaps 1;
QY      1336 ATGTTTGACAGCTTATCATCGATAAGCTTTAATGCGGTAGTTTATCACAGTTAAATTGCT 1395
DB      8 ATGTTTGACAGCTTAT-ATCGATAAGCTTTAATGCGGTAGTTTATCACAGTTAAATTGCT 66
QY      1396 AACGAGTCAGGACCGGTGTATGAATCTAACATGCGCTCATCGTCACTCTCGGCACdG 1455
DB      67 AACGAGTCAGGACCGGTATGAATCTAACATGCGCTCATCGTCACTCTCGGCACdG 126
QY      1456 TCACCTCGATGCTTAGGCATAGCGTTGGTTATCGCGGTACTGCGGGCCCTTTGCGGG 1515
DB      127 TCACCTCGATGCTTAGGCATAGCGTTGGTTATCGCGGTACTGCGGGCCCTTTGCGGG 186
QY      1516 ATATCGTCCATTCGACAGCATCGGCAGTCACTATGGGTGCTGCTAGCGGTATATGCT 1575
DB      187 ATATCGTCCATTCGACAGCATCGGCAGTCACTATGGGTGCTGCTAGCGGTATATGCT 246
QY      1576 TGATGCAATTTCTATGCGCACCGCTTCTCGGAGCACTGTCCGACCGCTTTGGCCGCCGCC 1635
DB      247 TGATGCAATTTCTATGCGCACCGCTTCTCGGAGCACTGTCCGACCGCTTTGGCCGCCGCC 306
QY      1636 CAGTCTCTGCTCGCTTCTGCTACTTTGAGGCACATATCGACTACGCGATCATGGCGACACAC 1695
DB      307 CAGTCTCTGCTCGCTTCTGCTACTTTGAGGCACATATCGACTACGCGATCATGGCGACACAC 366
QY      1696 CCGTCTCTGGATCTCTACGCGCGAGCAGCATCGTGGCCGCGCATCCGCGCGCCACAGGTG 1755

```

```

DB      367 CCGTCCTGTGGATCCTCTACGCGGACGATCGTGGCCGGCATCACCGGGGCCACAGGTG 426
QY      1756 CGGTTGCTGGCGCTATATCGCGACATCACCGATGGGAAGATCGGGCTCGCCACTTCG 1815
DB      427 CGGTTGCTGGCGCTATATCGCGACATCACCGATGGGAAGATCGGGCTCGCCACTTCG 486
QY      1816 GGCTCATGAGCGCTTGTTCGGCGCTGGGTATGGTGGCAGGCCCGCTGGCCGGGGACTGT 1875
DB      487 GGCTCATGAGCGCTTGTTCGGCGCTGGGTATGGTGGCAGGCCCGCTGGCCGGGGACTGT 546
QY      1876 TGGGCGCCATCTCTTCATGCAATGCAATCTCTTGGCGGGCGGTGCTCAACGGCTCAACC 1935
DB      547 TGGGCGCCATCTCTTCATGCAATGCAATCTCTTGGCGGGCGGTGCTCAACGGCTCAACC 606
QY      1936 TACTACTGGGCTGCTTCTTAATGCGAGGTGCGATAGGAGGAGCGTGCAGCCGATGCCCT 1995
DB      607 TACTACTGGGCTGCTTCTTAATGCGAGGTGCGATAGGAGGAGCGTGCAGCCGATGCCCT 666
QY      1996 TGAGAGGCT 2004
DB      667 TGCAAGCTT 675

RESULT 14
BJ683711      676 bp      mRNA      linear      EST 25-JAN-2005
LOCUS        BJ683711 HCEST library Haplochromis chilotes cDNA clone no86a05,
DEFINITION   mRNA sequence.
ACCESSION    BJ683711
VERSION      BJ683711.1 GI:46526832
KEYWORDS     EST.
SOURCE       Haplochromis chilotes
ORGANISM     Haplochromis chilotes
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
              Acanthomorpha; Acanthopterygii; Percomorpha; Perciformes;
              Labroidae; Cichlidae; African cichlids; Pseudocrenilabrinae;
              Haplochromini; Haplochromis.
REFERENCE    1 (bases 1 to 676)
AUTHORS      Watanabe M., Kobayashi N., Shin-i, T., Horiike, T., Tateno, Y.,
              Kohara, Y. and Okada, N.
TITLE        Extensive analysis of ORF sequences from two different cichlid
              species in Lake Victoria provides molecular evidence for a recent
              radiation event of the Victoria species flock: identity of EST
              sequences between Haplochromis chilotes and Haplochromis sp.
JOURNAL      Gene 343 (2), 263-269 (2004)
PUBMED       15588581
COMMENT      Contact: Tadasu Shin-i
              Center For Genetic Resource Information
              National Institute of Genetics
              1111 Yata, Mishima, Shizuoka 411-8540, Japan
              Tel: 81-559-81-6856
              Fax: 81-559-81-6855
              Email: tshini@genes.nig.ac.jp.
FEATURES     source
              1..676
              /organism="Haplochromis chilotes"
              /mol_type="mRNA"
              /db_xref="taxon:257977"
              /clone="no86a05"
              /tissue_type="jaw"
              /dev_stage="varied"
              /clone_lib="HCEST library"
ORIGIN
Query Match      13.3%; Score 641; DB 3; Length 676;
Best Local Similarity 98.4%; Pred. No. 6.1e-168;
Matches 658; Conservative 0; Mismatches 10; Indels 1; Gaps 1;
QY      1336 ATGTTTGACAGCTTATCATCGATAAGCTTTAATGCGGTAGTTTATCACAGTTAAATTGCT 1395
DB      8 ATGTTTGACAGCTTAT-ATCGATAAGCTTTAATGCGGTAGTTTATCACAGTTAAATTGCT 66

```


| | | | | |
|----|------|------|---|------|
| Qy | 1396 | AA | CGCAGCTCAGGCACCGGTGATGAAATCTAACATGCGCTCATCGTCTCATCTCGGCACCG | 1455 |
| Db | 67 | AAC | GCAGTCAAGCAGCACCCTGTATGAAATCTAACATGCGCTCATCGTCTCATCTCGGCACCG | 126 |
| Qy | 1456 | TC | ACCTTGGATGCTGTAGGCATAGGCTTTGGTTATGCGCGTACTGCCGGGCGCTCTTTGCGGG | 1515 |
| Db | 127 | TC | ACCTTGGATGCTGTAGGCATAGGCTTTGGTTATGCGCGTACTGCCGGGCGCTCTTTGCGGG | 186 |
| Qy | 1516 | ATA | TGCTGCCATTCCGACGAGATGCCAGTCACTATGCGGTGCTGTAGCGCTATATCGCT | 1575 |
| Db | 187 | ATA | TGCTGCCATTCCGACGAGATGCCAGTCACTATGCGGTGCTGTAGCGCTATATCGCT | 246 |
| Qy | 1576 | TG | ATGCAATTTCTATGCGCACCCGTTCTCGGAGCACTGTCCGACCGCTTTGGCGCGCGCC | 1635 |
| Db | 247 | TG | ATGCAATTTCTATGCGCACCCGTTCTCGGAGCACTGTCCGACCGCTTTGGCGCGCGCC | 306 |
| Qy | 1636 | CAG | TCCTGCTCGCTTCGCTACTTTGGAGCCACTATCGACTACGGGATCATGGGACACAC | 1695 |
| Db | 307 | CAG | TCCTGCTCGCTTCGCTACTTTGGAGCCACTATCGACTACGGGATCATGGGACACAC | 366 |
| Qy | 1696 | CCG | TCCTGTGGATCCTCTACGCCGGAAGCATCGTGGCGCGCATCACCGCGGCCACAGGTG | 1755 |
| Db | 367 | CCG | TCCTGTGGATCCTCTACGCCGGAAGCATCGTGGCGCGCATCACCGCGGCCACAGGTG | 426 |
| Qy | 1756 | CGG | TGCTGGCGCCTATATCCGCGACATACCGATGGGGAAGATCGGGCTCGCCACTTCCG | 1815 |
| Db | 427 | CGG | TGCTGGCGCCTATATCCGCGACATACCGATGGGGAAGATCGGGCTCGCCACTTCCG | 486 |
| Qy | 1816 | GGC | TCATAGCGCTTGTTCGGCGTGGGTATGGTGGCAGCCCGCTGGCGGGGACTGT | 1875 |
| Db | 487 | GGC | TCATAGCGCTTGTTCGGCGTGGGTATGGTGGCAGCCCGCTGGCGGGGACTGT | 546 |
| Qy | 1876 | TGG | CGCCATCTCTCTTGCATGCAACCATTCCTTGGCGCGCGGTGCTCAAAGGCGCTCAAAC | 1935 |
| Db | 547 | TGG | CGCGCATCTCTCTTGCATGCAACCATTCCTTGGCGCGCGGTGCTCAAAGGCGCTCAAAC | 606 |
| Qy | 1936 | TACT | TACTGGGCTGCTTCTTAATGACAGATCGCATAGAGGAGAGCGCTCGACCGATGCGCT | 1995 |
| Db | 607 | TACT | TACTGGGCTGCTTCTTAATGACAGATCGCATAGAGGAGAGCGCTCGACCGATGCGCT | 666 |
| Qy | 1996 | TG | AGAGCCT 2004 | |
| Db | 667 | TG | CAAGCTT 675 | |

| | |
|------------|---|
| RESULT | 15 |
| LOCUS | BJ683533 linear EST 25-JAN-2005 |
| DEFINITION | BJ683533 HCBST library Haplochromis chilotes cDNA clone no82a09, mRNA sequence. |
| ACCESSION | BJ683533 |
| VERSION | BJ683533.1 GI:46526654 |
| KEYWORDS | EST. |
| SOURCE | Haplochromis chilotes Haplochromis chilotes |
| ORGANISM | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorphia; Perciformes; Labroidae; Cichlidae; African cichlids; Pseudocrenilabrinae; Haplochromini; Haplochromis. 1 (bases 1 to 683) |
| REFERENCE | Watanabe,M., Kobayashi,N., Shin-i,T., Horike,T., Tateno,Y., Kohara,Y. and Okada,N. |
| AUTHORS | |
| TITLE | Extensive analysis of ORF sequences from two different cichlid species in Lake Victoria provides molecular evidence for a recent radiation event of the Victoria species flock; identity of EST sequences between Haplochromis chilotes and Haplochromis sp. |
| JOURNAL | Gene 343 (2), 263-269 (2004) |
| PUBMED | 15588581 |
| CONTACT | Contact: Tadasu Shin-i Center for Genetic Resource Information National Institute of Genetics |

| FEATURES | | source | | ORIGIN | |
|--|--|-----------------------------------|--|----------------------------------|--|
| 1111 Yata, Mishima, Shizuoka 411-8540, Japan Tel: 81-559-81-6856 Fax: 81-559-81-6855 Email: tshini@genes.nig.ac.jp. | | Location/Qualifiers | | | |
| 1..683 | | /organism="Haplochromis chilotes" | | | |
| | | /mol_type="mRNA" | | | |
| | | /db_xref="taxon:257977" | | | |
| | | /clone="nc82a09" | | | |
| | | /tissue_type="jaw" | | | |
| | | /dev_stage="varied" | | | |
| | | /clone_lib="HCESr library" | | | |
| Query Match | | 13.2%; | | Score 637.4; DB 3; Length 683; | |
| Best Local Similarity | | 98.3%; | | Pred. No. 6.3e-167; | |
| Matches 644; | | Conservative 0; | | Mismatches 11; Indels 0; Gaps 0; | |
| Qy | 1353 ATCGATAAGCTTTAAATGCGGTAGTGTATACACAGTTAAATGCTTAACGCGAGTCAGGCACCG | 1412 | | | |
| Db | 24 ATCGATAAGCTTTAAATGCGGTAGTGTATACACAGTTAAATGCTTAACGCGAGTCAGGCACCG | 83 | | | |
| Qy | 1413 TGTATGAATCTAAACAATGCGTTCATCGTCATCCTCGGCACCGTCACCCCTGGATGCTGTA | 1472 | | | |
| Db | 84 TGTATGAATCTAAACAATGCGTTCATCGTCATCCTCGGCACCGTCACCCCTGGATGCTGTA | 143 | | | |
| Qy | 1473 GGCATAGGCTTGTTATGCGGTACTCGCGGGCTCTTGGGGGATATCGTCCATTCGCAC | 1532 | | | |
| Db | 144 GGCATAGGCTTGTTATGCGGTACTCGCGGGCTCTTGGGGGATATCGTCCATTCGCAC | 203 | | | |
| Qy | 1533 AGCATGCCAGTCACATATGCGGTGCTAGCGCTATATGCGTTGATGCAATTCCTATGC | 1592 | | | |
| Db | 204 AGCATGCCAGTCACATATGCGGTGCTAGCGCTATATGCGTTGATGCAATTCCTATGC | 263 | | | |
| Qy | 1593 GCACCCGTTCTCGAGCAGCTGTCCGACCGCTTTGGCGCGCGCCACGTCCTGCTCGCTTCG | 1652 | | | |
| Db | 264 GCACCCGTTCTCGAGCAGCTGTCCGACCGCTTTGGCGCGCGCCACGTCCTGCTCGCTTCG | 323 | | | |
| Qy | 1653 CTACTTTGGAGCCACTATCGACTACGGCGACATCGGCACACACCCGCTCTGTGGATCCCTC | 1712 | | | |
| Db | 324 CTACTTTGGAGCCACTATCGACTACGGCGACATCGGCACACACCCGCTCTGTGGATCCCTC | 383 | | | |
| Qy | 1713 TAGCGCGAGCGCATCGTGGCGGCGCATACGGCGCCACAGGTGCGGTTCGTGGCGGCTAT | 1772 | | | |
| Db | 384 TAGCGCGAGCGCATCGTGGCGGCGCATACGGCGCCACAGGTGCGGTTCGTGGCGGCTAT | 443 | | | |
| Qy | 1773 ATCGCGGACATCACCGATGGGGAAGATCGGGCTCGCCACTTCGGGCTCATAGAGCGCTTGT | 1832 | | | |
| Db | 444 ATCGCGGACATCACCGATGGGGAAGATCGGGCTCGCCACTTCGGGCTCATAGAGCGCTTGT | 503 | | | |
| Qy | 1833 TTCCGCGTGGGTATGTTGGCAGGCCCCGCTGGCGGGGAGCTGTGGGGCGGCATCTCCTTTCG | 1892 | | | |
| Db | 504 TTCCGCGTGGGTATGTTGGCAGGCCCCGCTGGCGGGGAGCTGTGGGGCGGCATCTCCTTCG | 563 | | | |
| Qy | 1893 CATGCACCATTCCTTCGCGCGCGGTGCTCAACGGGCTCAACCTACTACTGCGGCTGCTTC | 1952 | | | |
| Db | 564 CATGCACCATTCCTTCGCGCGCGGTGCTCAACGGGCTCAACCTACTACTGCGGCTGCTTC | 623 | | | |
| Qy | 1953 CTAATGAGAGTTCGATTAAGGAGAGCGTCGACCGATGCCCTTTGAGAGCCTTCA | 2007 | | | |
| Db | 624 CTAATGAGAGTTCGATTAAGGAGAGCGTCGACCGATGCCCTTTGAGAGTTCGAC | 678 | | | |

Search completed: January 18, 2006, 11:40:55
Job time : 20380.3 secs

THIS PAGE BLANK (USPTO)